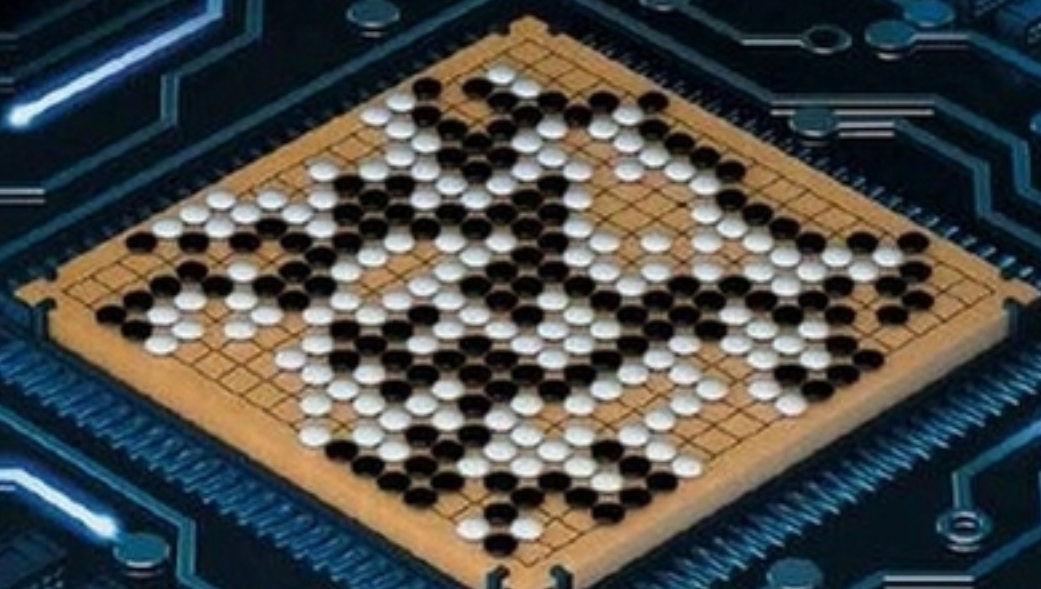


nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



At last — a computer program that
can beat a champion Go player **PAGE 484**

ALL SYSTEMS GO

CONSERVATION

SONGBIRDS À LA CARTE

Illegal harvest of millions
of Mediterranean birds

PAGE 452

RESEARCH ETHICS

SAFEGUARD TRANSPARENCY

Don't let openness backfire
on individuals

PAGE 459

POPULAR SCIENCE

WHEN GENES GOT 'SELFISH'

Dawkins's calling
card 40 years on

PAGE 482

NATURE.COM/NATURE

28 January 2016 £10

Vol. 529, No. 7587

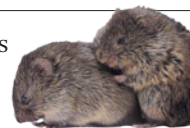


THIS WEEK

EDITORIALS

PSYCHOLOGY Don't believe the internal 'imposter' voices **p.438**

WORLD VIEW India needs a home-grown attitude on GM **p.439**



BEHAVIOUR Voles help stressed peers out of holes **p.441**

Digital intuition

A computer program that can outplay humans in the abstract game of Go will redefine our relationship with machines.

Napoleon had it and so did Charles Darwin. Tennis champion Roger Federer has it in spades. The dictionary defines intuition as knowledge obtained without conscious reasoning. It is decision-making based on apparently instinctual responses; thinking without thinking.

Intuition is a very human skill, or so we like to think. Or, more accurately, so we liked to think. In what could prove to be a landmark moment for artificial intelligence, scientists announce this week that they have created an intuitive computer. The machine acts according to its programming, but it also chooses what to do on the basis of something — knowledge, experience or a combination of the two — that its programmers cannot predict or fully explain. And, in the limited tests carried out so far, the computer has proved that it can make these intuitive decisions much more effectively than the most skilled humans can. The machines are not just on the rise, they have nudged ahead.

Experts in ethics, computer science and artificial intelligence routinely debate whether clever machines in the future will use their powers for good or evil. This latest example of digital discovery puts neural networks to work on a problem that is almost as old: how to win at the board game Go.

Outside business-management seminars, Go is not well known in the West, but it is older, more complex and harder to master than chess. Yet it is simpler to learn and play: two players take it in turns to place black or white counters on a grid. When a counter (called a stone) is surrounded by rivals, it is removed from the board. Winning — like so much in life and war — is about controlling the most territory. The game is wildly popular across countries in east Asia, and players from Japan, China and South Korea routinely compete in televised professional tournaments.

Computers mastered chess two decades ago, when IBM's Deep Blue machine won against then-world-champion Garry Kasparov in 1997, but Go was thought to be safe from artificial conquest. That is partly because all of the possible moves in Go, as well as the resulting combinations of stones on the board, are much too numerous for any computer to crunch through and compare to select one manoeuvre. (The same goes for chess, but the diversity in the value of chess pieces enables some short cuts.) In Go, all stones are worth the same and their influences can be felt through vast distances across the board.

On page 484 of this issue, computer scientists at Google DeepMind in London unveil the successor to Deep Blue. It is a program called AlphaGo, and in October 2015 it beat the human Go champion of Europe by five games to zero. To put that into context, in Deep Blue's time, a human beginner with just a week's practice could easily defeat the best Go computer programs. A match between AlphaGo and the world's most titled player of the decade is lined up for March (see page 445).

AlphaGo cannot explain how it chooses its moves, but its programmers are more open than Deep Blue's in publishing how it is built. Previous Go computer programs explore moves at random,

but the new technology relies on a suite of deep neural networks. These were trained to mimic the moves of the best human players, to reward wins and, using a probability distribution, to limit the outcomes for any board position to a single verdict: win or lose. Working together, these machine-learning strategies can massively reduce the number of possible moves the program evaluates and chooses from — in a seemingly intuitive way.

"The machine becomes an oracle; its pronouncements have to be believed."

As shown by its results, the moves that AlphaGo selects are invariably correct. But the interplay of its neural networks means that a human can hardly check its working, or verify its decisions before they are followed through. As the use of deep neural network systems spreads into everyday life — they are already used to analyse and recommend financial transactions — it raises an interesting concept for humans and their relationships with machines. The machine becomes an oracle; its pronouncements have to be believed.

When a conventional computer tells an engineer to place a rivet or a weld in a specific place on an aircraft wing, the engineer — if he or she wishes — can lift the machine's lid and examine the assumptions and calculations inside. That is why the rest of us are happy to fly. Intuitive machines will need more than trust: they will demand faith. ■

In praise of parks

Our affection for national parks is well founded, but many more areas need protection.

Yellowstone, the world's first national park, was created in 1872. It took rather longer for politicians to set up an agency to actually oversee such places: they got around to that in 1916. So the US National Park Service celebrates its centenary this year.

The agency also marked a shift in the way people think about parks. Yellowstone, which lies mostly in Wyoming, has little in common with the manicured gardens enjoyed by European gentry or admired by ancient Chinese kings. It and other huge, wild national parks are places where nature can supposedly be seen unmodified and unadorned, far from the pollution and bustle of cities.

Like much contemporary thinking, this rather ignores the history of native peoples and their stewardship of swathes of land before the arrival of Europeans. But this relatively new idea of parks as a wild refuge from the modern world has taken root. The United States' national parks have become some of the most iconic places in the

country. Yellowstone and Glacier National Park in Montana rival the White House and the Smithsonian as tourist attractions.

Similarly beloved national parks exist in other countries. The United Kingdom protected the Peak District in 1951 and now has 15 national parks. China started to protect nature reserves in the 1950s and now has jewels such as the Zhangjiajie National Forest Park and Jiuzhaigou nature reserve. In 2007, Pudacuo National Park became what is sometimes claimed to be the country's first true 'national park' (as it reaches standards laid down by the International Union for Conservation of Nature).

It has even been suggested that cities themselves can be parks, rather than just containing them. A campaign has been launched to have London declared a kind of urban national park. This might seem a backwards device — in general, parks are established in beautiful places that people love, not established to make places beautiful and encourage people to love them. But it goes to show the affection that many feel towards places classified as parks, be they vast national expanses or local patches of scrubby grass.

This affection is not based solely on a misty-eyed yearning for the outdoors. There is an evidence base that parks are a good thing. Many studies have confirmed that they come with significant benefits. They seem to make people who use them healthier and happier. They make local ecosystems more diverse and more resilient. They can even help to mitigate climate change to a small degree.

But not everyone is happy when land is set aside in parks and other uses are limited. In the United States, a group of armed men have seized — and, as *Nature* went to press, were still in control of — the Malheur National Wildlife Refuge in Oregon. Although there are a plethora of issues related to that act of insurgency, this event is linked to a dispute over attempts by the federal government to control cattle grazing so as to protect a species of tortoise.

This situation might be extreme. But the story of conflict between

park authorities and people who may once have worked inside park boundaries, or who wish to work there, is universal. Last week, 60 non-governmental organizations again raised the issue of threats to the Virunga National Park in the Democratic Republic of the Congo, one of the last remaining strongholds for mountain gorillas. The prospect of drilling for oil in the park itself has been of concern in the past, and environmental groups are now warning that oil drilling in nearby Uganda could harm the ecosystem of which the park forms a part.

“Setting aside an area as a park should not be used as a fig leaf for a lack of a wider environmental approach.”

Things have been equally fraught at sea. As governments have created more and more 'marine protected areas', fishermen have railed against being excluded from waters they once hauled nets in. Researchers have questioned whether many of these areas are actually protecting what needs to be safeguarded. And there are questions

about just how protected some of these areas are, and whether countries are gaming systems to hit international targets.

The spirit of international targets to protect 17% of terrestrial areas and 10% of marine areas certainly intends that they be reached by protecting places that warrant support, not those that are easy to protect because no one cares about exploiting what is there.

Paradoxically, as it becomes ever more apparent that we need to protect areas of outstanding beauty and delicate ecology, it is becoming increasingly clear that it is not enough to do only this. Setting aside an area as a park should not be used as a fig leaf for a lack of a wider environmental approach. Cities, agricultural landscapes, wasteland and seas open to industry all need to be managed in a sensible and planned fashion.

We need more parks. But the real challenge is to make people treat the whole planet with the respect that most show to their parks. ■

Found out

Self-doubt is a pernicious affliction that can overwhelm researchers.

Oh good grief, why did I ever say that I would write something about imposter syndrome? What do I know about it, really? I'm not a psychologist or a researcher or a proper expert, I'm just a journalist. I thought I knew what imposter syndrome was — that some people don't call it a syndrome as such, because that implies a mental disorder. And I thought that I had suffered from those feelings of doubt and inadequacy about my abilities, but now I'm not sure. Maybe other people just suffer from imposter syndrome more badly than I do.

What if I simply tell people to go and read the Careers feature on page 555 that describes how imposter syndrome can affect people in science, and which offers some useful tips on overcoming what, as it turns out, are very common feelings? But then again, won't that make it clear that I don't have anything else to say?

Maybe I can deflect attention from my own pitiful performance by citing talented celebrities who have admitted to sometimes feeling like frauds and imposters. The multiple-Oscar-winning film star Meryl Streep perhaps? I'm sure I read somewhere, though I might be wrong, that she once said she couldn't understand why anyone would want to watch her on screen because she felt she couldn't act. Or the famous and award-gathering author Maya Angelou, who after each of her eleven books, said she felt that this was the time she was going to be found out.

See, I have done the research. I do know what I am talking about,

so why does it feel as if everyone around me is simply better at this than me? I bet that's the way the editor thinks, too. Maybe this would be a good time to throw in an Einstein quote, and seek some reflected glory: “The exaggerated esteem in which my lifework is held makes me very ill at ease. I feel compelled to think of myself as an involuntary swindler.”

I wish I had that Dunning-Kruger effect, the almost opposite experience to imposter syndrome in which people who really aren't qualified or knowledgeable show remarkable (and misplaced) confidence in their abilities and decisions. Life would be so much easier then, or at least it would seem that way.

The thing about imposter syndrome is that it's been known and written about since the late 1980s, and yet each generation of young scientists (and teachers, nurses, jet pilots and so on) feel isolated and anxious because of it. They feel that they are the only ones to have these crippling self-doubts, as if someone is about to tap them on the shoulder and confess that the whole situation — the job, the responsibility, the career — is an elaborate hoax and they should go home and stop being so presumptuous as to believe that they had anything to offer.

They need to know that these thoughts and ideas are common, and in fact are most common among genuine high achievers. They should be told that rejection — of papers, grants, ideas — in science is the norm and that they shouldn't lose heart when it happens. After all, this is a field of human endeavour in which experts boast about how little they know and proudly display their margins of error. Young and vulnerable researchers need to know that if they tell someone — a

friend or colleague or mentor — about how they are feeling, then they will almost certainly hear the words 'me too' and will feel better.

I should tell them that. If only I could find the right words. ■

➔ **NATURE.COM**

To comment online,
click on Editorials at:
go.nature.com/xhunq



India needs home-grown GM food to stop starvation

Indian scientists must develop domestic genetically modified crops rather than rely on unsuitable foreign technology, says Anurag Chaurasia.

At the beginning of this month, Prime Minister Narendra Modi announced a road map to guide India's science and technology over the next two decades. Launched during the Indian Science Congress at the University of Mysore, the plan signalled a cautious approach to techniques such as genetically modified (GM) crops, noting that "some aspects of biotechnology have posed serious legal and ethical problems in recent years". That is true, but a different and much larger problem looms for India. According to the 2015 United Nations World Population Prospects report, India will surpass China by early next decade as the most populous country on Earth, with the most mouths to feed. India is already classed as having a 'serious' hunger problem, according to the 2015 Global Hunger Index of the International Food Policy Research Institute. There is a danger that many of these new Indians will not have sufficient food.

Where can additional food come from? Grain production is stagnant, and rapid urbanization is reducing available land. To increase food production, India needs to invest in modern agricultural methods, including GM crops.

Indian researchers have shown that they have the expertise to generate GM plants, most obviously the pest-resistant cotton that is now widely grown in India. But almost all of this work has relied on molecular-biology research done elsewhere — India has in effect borrowed or been given the genes. This leads to complications, usually conflict over intellectual property (IP) rights.

Most high-profile was the insecticide-producing GM cotton variety that was released by the Indian Council of Agricultural Research in New Delhi in 2009. It was based on a *Bacillus thuringiensis* gene to which the agricultural biotechnology company Monsanto, based in St Louis, Missouri, owed the IP rights. The ensuing controversy has seeded confusion among Indian researchers, scientific managers and administrators over IP rights, patents and the related rules and regulations.

In response, India is turning to research based on old discoveries, including genes that are in the public domain or no longer protected by patents. The problem here is that insects have already developed resistance to the toxins produced by such genes: the companies that developed first-generation GM crops with these genes are already on second- and even third-generation versions of the same plants. Increased use of this old technology in India can only accelerate resistance and make the situation more difficult. Other developing countries (including Pakistan) are also turning to such redundant technology.

India should stop trying to build the Taj Mahal with borrowed bricks. We need a concerted effort at home to discover and manipulate relevant

genes in indigenous organisms and crops (such as chickpea and rice). Indian microbial institutes should take up projects in this direction, because most of the currently used genes for transgenic generation are of microbial origin. That requires a change in direction from an Indian GM-food strategy that has traditionally aimed at quick product development instead of careful assessment of the underlying science.

Such home-grown GM crops would also reduce reliance on transgenic technology produced by multinational companies, which is expensive and rarely optimized for the conditions of specific regions. Some GM crops designed abroad need more water than is usually available in some parts of India, for example, putting great stress on farmers.

Indian scientists need better training in IP issues, especially when our researchers join foreign collaborations to examine and exploit the molecular biology of our natural resources. Otherwise, Indian researchers may get the scientific credit for discoveries but fail to claim the right to commercialize the products developed.

Indian regulators should exert tighter controls on IP rights. At present, they focus only on the export of physical material, such as seeds and tissue. They need also to monitor, and make claims on, molecular information drawn from this material, down to the level of genes and promoter regions. According to the Food and Agriculture Organization of the UN, India is the largest donor of crop germplasm to the world. Without realizing its importance, we are giving away the rights to exploit one of our most precious assets.

Agrarian India is excelling in space science, but it needs to focus closer to home as well. It needs to follow the example of China, which is

slowly but steadily building a GM-food market that is based on domestic discoveries. Compared with China, India has three times as much land planted with GM crops, but whereas India's plants were mostly created with technology bought from abroad, China's fields contain crops that were developed, tested and commercialized by Chinese scientists. India does not have to reject the expertise of international companies, but it must do more to build knowledge and skills at home.

Mahatma Gandhi only wore clothes that he had woven himself. He gave India the slogan "from *swadeshi* to *swaraaj*", which means "be indigenous in order to self-rule". The Indian government should take this message on board when planning future investment in biotechnology. The theme of this month's science congress, after all, was "science and technology for indigenous development in India". Indigenous development needs indigenous research. ■

Anurag Chaurasia is a biotechnologist with the National Bureau of Agriculturally Important Microorganisms in Kushmaur, India.
e-mail: govtofindia.icar@gmail.com

**SOME GM CROPS
DESIGNED
ABROAD
NEED MORE WATER
THAN IS USUALLY
AVAILABLE
IN SOME PARTS
OF INDIA.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/vxi84m

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

INFECTIOUS DISEASE

Antibody for range of ebolaviruses

Antibodies that recognize multiple ebolavirus species could treat the deadly infection.

Humans infected with ebolaviruses make antibodies that bind to proteins on the surface of virus particles. This prevents the virus from infecting more cells, but it is unclear whether antibodies for one of the species of ebolavirus will work for the others. To find out, Alexander Bukreyev at the University of Texas Medical Branch in Galveston, James Crowe at Vanderbilt University in Nashville, Tennessee, and their team cultured antibody-producing B cells from seven people who survived a 2007 Ebola outbreak in Uganda, caused by an ebolavirus species called Bundibugyo.

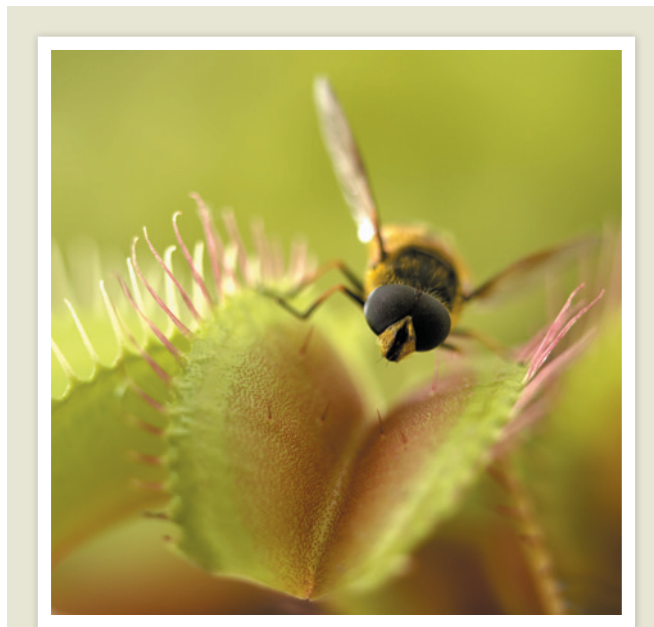
Some of the antibodies blocked two other species of ebolaviruses. One of the Bundibugyo antibodies protected guinea pigs that had been infected with another species of ebolavirus. **Cell** <http://doi.org/bb3g> (2016)

ASTRONOMY

Turbulence roils luminous galaxy

The brightest-known galaxy is blasting gas out into space — and providing astronomers with a rare glimpse of how extreme galaxies evolve.

Known as W2246-0526, the galaxy is as bright as 350 trillion Suns and is powered by a supermassive black hole at its heart. A team led by Tanio Díaz-Santos at Diego Portales University in Santiago, Chile, used



PLANT SCIENCE

Plants count to five

Venus flytraps count the number of touches from trapped insect prey before producing digestive juices.

Erwin Neher at the Max Planck Institute for Biophysical Chemistry in Göttingen and Rainer Hedrich of the University of Würzburg, both in Germany, and their colleagues touched the leaves of Venus flytraps (*Dionaea muscipula*; pictured) to mimic captured, moving prey. They recorded the plant's electrical impulses in response to 1–60 touches, and found that two impulses triggered the trap to close. But only after the fifth impulse did plants begin to synthesize the digestive enzyme hydrolase and increase their production of a sodium transporter, which is used to absorb nutrients.

Venus flytraps may record the touches of potential prey to identify insects that are worth digesting, the authors say. **Curr. Biol.** <http://doi.org/bbzb> (2016)

the high-resolution Atacama Large Millimeter/submillimeter Array in Chile to study carbon ions rushing outwards from the galaxy.

The gas races out at speeds of about 2 million kilometres an hour, violently illuminating the surrounding space. W2246-0526 might be spewing out much of its energy, and could become more tame in the future.

Astrophys. J. Lett. 816, L6 (2016)

BIOENGINEERING

Shielded cells treat diabetes

Insulin-producing cells derived from human stem cells restore blood sugar to normal levels when encased in a porous biomaterial and implanted in diabetic mice.

People with severe type 1 diabetes can sometimes be treated with a transplant of

insulin-producing cells from cadavers, but the cell supply is limited. The recipient must also stay on drugs to stop the immune system attacking the cells. Instead, Daniel Anderson of the Massachusetts Institute of Technology in Cambridge and his colleagues derived insulin-producing β -cells from human embryonic stem cells and encapsulated them in a substance called TMTD alginate.

When implanted into diabetic mice, the coated cells were shielded from immune attack. The animals also maintained normal blood-sugar control until the implants were removed 174 days later.

Nature Med. <http://dx.doi.org/10.1038/nm.4030> (2016)

OCEANOGRAPHY

Rising seas differ by region

The expansion of oceans as the climate warms has contributed to a rise in global sea levels of about 1.38 millimetres per year — roughly twice that of previous estimates.

Roelof Rietbroek at the University of Bonn in Germany and his colleagues analysed sea-surface heights from satellite radar data, and looked at changes in water storage from the Gravity Recovery And Climate Experiment (GRACE). They found that between 2002 and 2014, sea levels have increased in total by roughly 2.74 mm per year, with 1.38 mm of that coming from ocean thermal expansion and 1.08 mm from melting ice sheets, glaciers and other water sources on land, such as rivers.

The team also uncovered large regional differences. For example, the Philippines experienced a sea-level rise of about 14.7 mm per year, mostly because of thermal

237/ADAM GAULT/OCEANO/CORBIS

expansion of the ocean, whereas the central and eastern Pacific saw decreases. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1519132113> (2016)

METABOLISM

Beige fat boosts metabolism

Human 'beige' fat cells implanted in mice can improve the animals' glucose metabolism and liver-fat profiles.

The presence of beige fat — brown fat cells within white fat-storing tissue — is correlated with better metabolic health, but it was not known whether beige fat causes this. To see whether there is a causal link, Silvia Corvera of the University of Massachusetts Medical School in Worcester and her colleagues grew human beige fat cells in the lab, placed them in mice, and found that they formed well-defined adipose tissue. Animals with the implants had lower blood-glucose levels, absorbed the glucose more quickly than did untreated controls, and had less fat in their livers.

The results suggest that beige fat could have therapeutic use, the authors say.

Nature Med. <http://dx.doi.org/10.1038/nm.4031> (2016)

MATERIALS

Add water for 3D-printed flowers

Researchers have 3D-printed hydrogel composites that swell and morph into flower shapes when immersed in water.

Lakshminarayanan

Mahadevan and Jennifer Lewis at Harvard University in Cambridge, Massachusetts, and their colleagues used an ink made of cellulose fibrils embedded in a hydrogel matrix, which mimics plant-cell walls and swells in water. By controlling the alignment of the fibrils in the ink during printing, the team produced flat materials that bend and twist when placed in water, producing structures that mimic flowers (**pictured**).

The approach could be used to create designer, shape-changing structures for biomedical applications or smart textiles, the authors say. *Nature Mater.* <http://dx.doi.org/10.1038/nmat4544> (2016)

CHEMISTRY

Polymers woven into stretchy web

Organic polymers woven into a 3D framework offer a new way of making flexible materials with tunable properties.

Covalent organic frameworks are highly porous structures with many promising applications, but they are typically rigid. Omar Yaghi of the University of California, Berkeley, Osamu Terasaki of Stockholm University and their colleagues created such a framework, dubbed COF-505. It is made of individual building blocks of copper ions that carry fragments of a polymer. Joining these units together with linear molecules formed crystals with the same tetrahedral geometry as diamond.

The researchers then removed the copper ions

SOCIAL SELECTION

Popular topics on social media

House bugs crawl over social media

Many commenters on Twitter this week felt their skin crawl after reading that some US households are home to more than 200 different species of insects and other creatures, according to one study. Entomologists collected more than 10,000 specimens of arthropods (insects and other animals with exoskeletons and segmented bodies) from 50 homes in Raleigh, North Carolina, and found surprising diversity. Their results, published in *PeerJ*, suggest that the average home contained 93 different species, from spiders and flies to cockroaches and beetles. Out of the 304 arthropod families identified, 149 were rare. And only 5 out of the 554 rooms examined — 4 bathrooms and 1 bedroom — contained no bugs at all. Joachim Maes, an ecologist at the European Commission's Institute for Environment and Sustainability in Ispra, Italy, tweeted: "We are literally surrounded by biodiversity." The study analysed only the types of species present, and the authors recommend a more-in-depth study

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/cchnu3

of confined spaces in homes — such as under the stairs — to get more-accurate data on the number and diversity of household bugs.

PeerJ 4, e1582 (2016)

to leave interwoven, helical polymer threads that were collectively ten times more elastic than the precursor. The copper ions could also be replaced, raising the possibility of loading the polymer weave with metal catalysts, or of using it to absorb metal ions from liquid waste.

Science 351, 365–369 (2016)

ANIMAL BEHAVIOUR

Voles console stressed friends

Prairie voles seem to console their distraught cage-mates — a behaviour previously seen only in humans and in other animals with advanced cognition, such as great apes and elephants.

James Burkett, Larry Young and their colleagues at Emory University in Atlanta, Georgia, separated pairs of prairie voles (*Microtus ochrogaster*; **pictured**) in the lab and measured how long the rodents groomed each other when they were reunited. Voies spent significantly more time grooming partners that



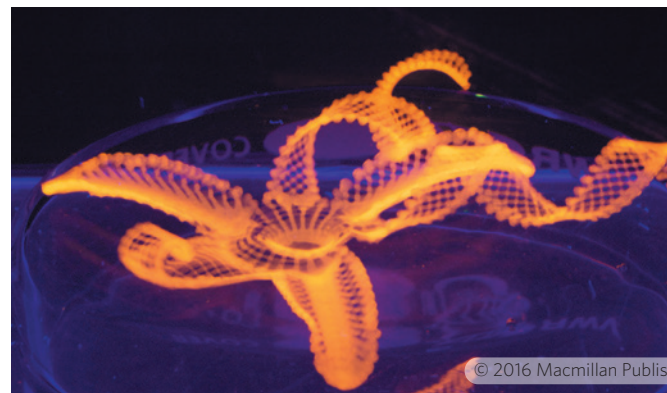
had been subjected to noise and mild electric shocks during the separation period, even though they had not observed the stressful event.

The unstressed voles showed the same levels of stress hormones as their stressed cage-mates. This response disappeared when the researchers chemically blocked the brain receptor for oxytocin, a hormone involved in empathy in humans.

Further research on this consolation behaviour in rodents could yield insight into certain psychiatric disorders that involve a lack of empathy, the authors say.

Science 351, 375–378 (2016)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch



SEVEN DAYS

The news in brief

EVENTS

Hottest year

Last year was the hottest on record, according to data released on 20 January by NASA, the US National Oceanic and Atmospheric Administration (NOAA) and the UK Met Office. All three organizations document unprecedented high temperatures in 2015, pushing the global average to more than 1 °C above pre-industrial levels. A powerful El Niño weather system, marked by warmed waters in the tropical Pacific Ocean, helped to drive atmospheric temperatures well above those for 2014, the previous record high. Some researchers suggest that broader trends in the Pacific could spell even more dramatic temperature increases in years to come. See page 450 for more.

Davos debates

Health, science and environmental issues were high on the agenda at the World Economic Forum (WEF) meeting in Davos, Switzerland, where international leaders met on 20–23 January. The 2016 meeting — which had as its theme ‘the fourth industrial revolution’ — saw 85 pharmaceutical companies demand new business models to incentivize the development of antibiotics to combat

NUMBER CRUNCH

274,207,281–1

The largest known prime number, at more than 22 million digits. The find was announced this month by the Great Internet Mersenne Prime Search (www.mersenne.org).



S. D. BIJU

DNA search reveals new frog genus

DNA analyses have unveiled a new genus of frogs. The group, *Frankixalus*, hails from northeastern India and includes a taxonomically challenging amphibian, first described in 1876 and named *Polypedates jerdonii*. The species, now renamed *Frankixalus jerdonii* (pictured), is joined in the new genus by one other — currently

unnamed — species. Classification of these tree-hole dwelling frogs had been confounded by their large snout vents and the high degree of webbing between their toes. But a genetic analysis led by S. D. Biju at the University of Delhi places the frogs in a distinct group (S. D. Biju *et al.* *PLoS ONE* 11, e0145727; 2016).

drug-resistant infections. Environmental discussions included a WEF report predicting that the oceans would contain more plastic than fish by weight by 2050, and an updated Environmental Protection Index report from Yale University in New Haven, Connecticut, which shows an increased degradation of air quality around the world.

Turkish protest

An international coalition of 20 higher-education organizations has written to the Turkish president, Recep Tayyip Erdoğan, urging him to reaffirm Turkey's commitment to academic freedom and freedom of expression. The 21 January letter, from

organizations including the European University Association and the Scholars at Risk Network, voices concern about academics who signed a petition calling on the government to end violent conflict in the country's predominantly Kurdish southeast. Erdoğan had accused the petition's signatories of spreading terrorist propaganda, and the letter says that 1,128 academics have since been placed under investigation.

Zika virus fears

Concern over the mosquito-borne Zika virus outbreak in the Americas is growing. In a statement on 24 January, the Pan American Health

Organization warned that Zika is likely to reach all countries in the region where the *Aedes* mosquitoes that transmit the virus live. The organization, a regional office of the World Health Organization, called for improved mosquito control, and advised people, especially pregnant women, to protect themselves from bites. In a risk assessment on 21 January, the European Centre for Disease Prevention and Control called for more surveillance of travellers returning from affected countries, so as to rapidly identify potential cases. This followed a 15 January travel warning from the US Centers for Disease Control and Prevention about Zika, which

some evidence has linked to babies born in Brazil with unusually small heads and brains.

POLICY

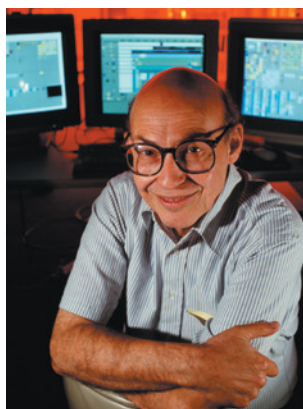
Data-sharing rule

The International Committee of Medical Journal Editors has proposed that it should be mandatory for research-paper authors to share the data underlying their findings. In an editorial published in several prestigious medical journals on 20 January, the group said that authors should share anonymized patient data with other researchers no later than six months after publication (see, for example, D. B. Taichman *Ann. Intern. Med.* <http://doi.org/bb4h>; 2016). The committee also said that a plan for how data will be shared should be part of the clinical-trial registration process. Signatories to the editorial include the editors-in-chief of *The New England Journal of Medicine* and *The Journal of the American Medical Association*.

PEOPLE

AI pioneer dies

Artificial intelligence (AI) expert Marvin Minsky (pictured) died, aged 88, on 24 January. After gaining his PhD at Princeton University in New Jersey, Minsky joined



the Massachusetts Institute of Technology (MIT) in Cambridge, where he co-founded the AI Laboratory and the MIT Media Lab. He built the world's first neural-network simulator, invented the confocal microscope and received almost all of the major awards in his field.

Berkeley lab head

Particle physicist Michael Witherell will become director of the Lawrence Berkeley National Laboratory on 1 March. The University of California, which manages the lab on behalf of the US Department of Energy, announced the move on 21 January. Located in Berkeley, California, the lab has an annual budget of nearly US\$800 million and is home to research in fields ranging from materials science to genomics.

Witherell is currently vice chancellor for research at the University of California, Santa Barbara, and led the Fermi National Accelerator Laboratory in Batavia, Illinois, from 1999 to 2005.

Koch quits

Billionaire David Koch has left his long-standing position on the board of trustees of the American Museum of Natural History in New York City, *The New York Times* reported on 20 January. Koch has given millions of dollars to the museum, which named its dinosaur wing after him. But his presence on the board has been criticized by scientists and activists because the Koch family has also funded climate-change deniers. Koch's last day on the board was 9 December. Koch remains on the advisory board of the Smithsonian National Museum of Natural History in Washington DC.

Freedom plea

More than 360 researchers from around the world have asked Iran to free retired polymer scientist Mohammad Hossein Rafiee-Fanoood, in a letter to the country's president, Hassan Rouhani. Rafiee, who is 71 years old, was imprisoned for "peaceful political activism", according to Amnesty International, which regards him as a prisoner

COMING UP

1 FEBRUARY

The Genome Integrity Discussion Group meets at the New York Academy of Sciences.

go.nature.com/crhom3

1–4 FEBRUARY

The Delhi Sustainable Development Summit is hosted by TERI, the energy and resources institute.

go.nature.com/ndvmaj

1–5 FEBRUARY

The Mares Conference on Marine Ecosystems Health and Conservation convenes in Olhão, Portugal.

www.maresconference.eu

of conscience. The letter, published on 26 January (see go.nature.com/7wflr7), says that he is kept in inhumane conditions and that the scientist's imprisonment breaches both the Constitution of the Islamic Republic of Iran and the International Covenant on Civil and Political Rights.

BUSINESS

Trade theft

Federal prosecutors in Pennsylvania have charged five people, including two researchers, with the theft of trade secrets from pharmaceutical giant GlaxoSmithKline (GSK). The researchers were employed at the company's facility in Upper Merion, Pennsylvania, and allegedly supplied confidential information about GSK's experimental therapies to a biotechnology company headquartered in China. The trade secrets in question include information about the production of therapeutic antibodies, including one in development to treat cancer.

NATURE.COM

For daily news updates see:

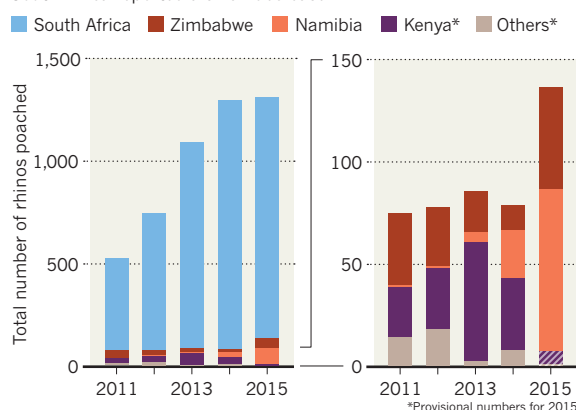
www.nature.com/news

TREND WATCH

At least 1,312 rhinoceros were illegally killed in Africa in 2015 — a record high for the continent, according to figures released on 21 January by TRAFFIC, a wildlife-trade monitoring network based in Cambridge, UK. South Africa, which has the largest share of this poaching, reported a small decrease — from 1,215 rhinos in 2014 to 1,175 in 2015. But that was more than offset by rises in illegal killing in neighbouring countries, according to TRAFFIC.

AFRICAN RHINO POACHING ON THE RISE

2015 is the worst year in decades for rhino poaching — although South Africa reported a small decrease.



NEWS IN FOCUS

CANINE PSYCHIATRY Dogs provide genetic clues to human disorders **p.446**

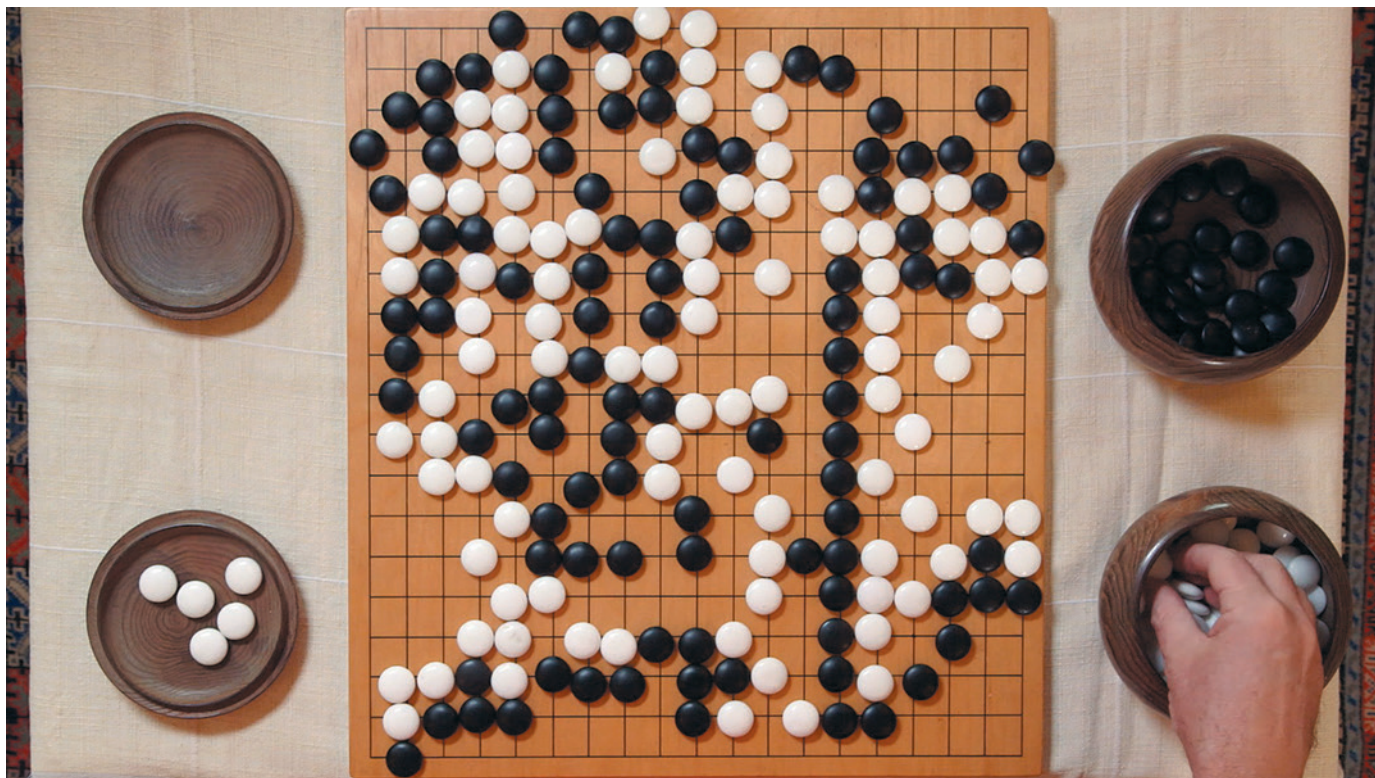
PHYSICS Debate over meaning of Stephen Hawking's latest paper **p.448**

CLIMATE Developing nations struggle to keep carbon accounts **p.450**

CONSERVATION Songbird killing for restaurants becomes a hot issue in Cyprus **p.452**



NATURE VIDEO



Go, a complex game popular in Asia, has frustrated the efforts of artificial-intelligence researchers for decades.

ARTIFICIAL INTELLIGENCE

Google masters Go

Deep-learning software excels at complex ancient board game.

BY ELIZABETH GIBNEY

A computer has beaten a human professional for the first time at Go — an ancient board game that has long been viewed as one of the greatest challenges for artificial intelligence (AI).

The best human players of chess, draughts and backgammon have all been outplayed by computers. But a hefty handicap was needed for computers to win at Go. Now Google's London-based AI company, DeepMind, claims that its machine has mastered the game.

DeepMind's program AlphaGo beat Fan Hui, the European Go champion, five times out of five in tournament conditions, the firm

reveals in research published in *Nature* on 27 January¹. It also defeated its silicon-based rivals, winning 99.8% of games against the current best programs. The program has yet to play the Go equivalent of a world champion, but a match against South Korean professional Lee Sedol, considered by many to be the world's strongest player, is scheduled for March. "We're pretty confident," says DeepMind co-founder Demis Hassabis.

"This is a really big result, it's huge," says Rémi Coulom, a programmer in Lille, France, who designed a commercial Go program called Crazy Stone. He had thought computer mastery of the game was a decade away.

The IBM chess computer Deep Blue, which

famously beat grandmaster Garry Kasparov in 1997, was explicitly programmed to win at the game. But AlphaGo was not preprogrammed to play Go: rather, it learned using a general-purpose algorithm that allowed it to interpret the game's patterns, in a similar way to how a DeepMind program learned to play 49 different arcade games².

This means that similar techniques could be applied to other AI domains that require recognition of complex patterns, long-term planning and decision-making, says Hassabis. "A lot of the things we're trying to do in the world come under that rubric." Examples are using medical images to make diagnoses or treatment plans, and improving climate-change models. ▶

► In China, Japan and South Korea, Go is hugely popular and is even played by celebrity professionals. But the game has long interested AI researchers because of its complexity. The rules are relatively simple: the goal is to gain the most territory by placing and capturing black and white stones on a 19×19 grid. But the average 150-move game contains more possible board configurations — 10^{170} — than there are atoms in the Universe, so it can't be solved by algorithms that search exhaustively for the best move.

ABSTRACT STRATEGY

Chess is less complex than Go, but it still has too many possible configurations to solve by brute force alone. Instead, programs cut down their searches by looking a few turns ahead and judging which player would have the upper hand. In Go, recognizing winning and losing positions is much harder: stones have equal values and can have subtle impacts far across the board.

To interpret Go boards and to learn the best possible moves, the AlphaGo program applied deep learning in neural networks — brain-inspired programs in which connections between layers of simulated neurons are strengthened through examples and experience. It first studied 30 million positions from expert games, gleaning abstract information on the state of play from board data, much as

other programmes categorize images from pixels (see *Nature* **505**, 146–148; 2014). Then it played against itself across 50 computers, improving with each iteration, a technique known as reinforcement learning.

The software was already competitive with the leading commercial Go programs, which select the best move by scanning a sample of simulated future games. DeepMind then combined this search approach with the ability to pick moves and interpret Go boards — giving

“Deep learning is killing every problem in AI.”

AlphaGo a better idea of which strategies are likely to be successful. The technique is “phenomenal”, says Jonathan Schaeffer, a computer scientist at the University of Alberta in Edmonton, Canada, whose software Chinook solved³ draughts in 2007. Rather than follow the trend of the past 30 years of trying to crack games using computing power, DeepMind has reverted to mimicking human-like knowledge, albeit by training, rather than by being programmed, he says. The feat also shows the power of deep learning, which is going from success to success, says Coulom. “Deep learning is killing every problem in AI.”

AlphaGo plays in a human way, says Fan. “If no one told me, maybe I would think the player was a little strange, but a very strong player, a real person.” The program seems to

have developed a conservative (rather than aggressive) style, adds Toby Manning, a lifelong Go player who refereed the match.

Google's rival firm Facebook has also been working on software that uses machine learning to play Go. Its program, called *darkforest*, is still behind commercial state-of-the-art Go AI systems, according to a November preprint⁴.

Hassabis says that many challenges remain in DeepMind's goal of developing a generalized AI system. In particular, its programs cannot yet usefully transfer their learning about one system — such as Go — to new tasks; a feat that humans perform seamlessly. “We've no idea how to do that. Not yet,” Hassabis says.

Go players will be keen to use the software to improve their game, says Manning, although Hassabis says that DeepMind has yet to decide whether it will make a commercial version.

AlphaGo hasn't killed the joy of the game, Manning adds. Strap lines boasting that Go is a game that computers can't win will have to be changed, he says. “But just because some software has got to a strength that I can only dream of, it's not going to stop me playing.” ■

SEE EDITORIAL P.437

1. Silver, D. *et al. Nature* **529**, 484–489 (2016).
2. Mnih, V. *et al. Nature* **518**, 529–533 (2015).
3. Schaeffer, J. *et al. Science* **317**, 1518–1522 (2007).
4. Tian, Y. & Zhu, Y. Preprint at arXiv <http://arxiv.org/pdf/1511.06410.pdf> (2015).

GENOMICS

Dog DNA probed for clues to human psychiatric ills

Project will compare gene data to owners' assessments of how their companions behave.

BY HEIDI LEDFORD

Addie plays hard for an 11-year-old greater Swiss mountain dog — she will occasionally ignore her advanced years to hurl her 37-kilogram body at an unwitting house guest in greeting. But she carries a mysterious burden: when she was 18 months old, she started licking her front legs aggressively enough to wear off patches of fur and draw blood.

Addie has canine compulsive disorder — a condition that is thought to be similar to human obsessive-compulsive disorder (OCD). Canine compulsive disorder can cause dogs to chase their tails for hours on end, or to suck on a toy or body part so compulsively that it interferes with their eating or sleeping.

Addie may soon help researchers to

determine why some dogs are more prone to the disorder than others. Her owner, Marjie Alonso of Somerville, Massachusetts, has enrolled her in a project called Darwin's Dogs, which aims to compare information about the behaviour of thousands of dogs against the animals' DNA profiles. The hope is that genetic links will emerge to conditions such as canine compulsive disorder and canine cognitive dysfunction — a dog analogue of dementia and possibly Alzheimer's disease. The project organizers have enrolled 3,000 dogs so far, but hope to gather data from at least 5,000, and they expect to begin analysing DNA samples in March.

“It's very exciting, and in many ways it's way overdue,” says Clive Wynne, who studies canine behaviour at Arizona State University in Tempe.

Researchers have long struggled to find

genetic links to human psychiatric disorders by analysing DNA samples from thousands of people. Those efforts have in recent years met with some success in schizophrenia and depression. But for some conditions, including OCD, not a single robust genetic link has been sifted from the background noise of normal genetic variation.

Human studies are difficult in part because the species is so genetically diverse, says Wynne. Dogs, however, are more genetically homogeneous. Selected over thousands of years for particular characteristics, they display less genetic variation than do humans. Pure-bred dogs, in particular, have been rendered highly genetically consistent to achieve a homogenous appearance and behaviour.

Dogs also live side-by-side with humans, which some think can make them a better

model for human disorders than mice living in a laboratory cage.

These qualities have made dogs attractive targets for studies of analogues to human ailments, including epilepsy, cancer and various psychiatric disorders. Border collies, for example, may over-react to loud noises in a manner akin to people with anxiety disorders. Geneticist Elinor Karlsson of the University of Massachusetts Medical School in Amherst and her colleagues have studied canine compulsive disorder, a condition that is particularly common in certain breeds, including Doberman pinschers. Their studies in 150 dogs have found possible links to four genes that encode proteins that act in the brain (R. Tang *et al.* *Genome Biol.* 15, R25; 2014).

To expand on those results, Karlsson has decided to go big. Limiting her studies to specific breeds would make it easier to pick out some genetic links, but others might be missed. So Karlsson and her colleagues, including Jesse McClure, a former dog trainer for the US Marine Corps, decided to collect data from mongrels as well as pure-bred dogs and to crowdsource the data collection.

That focus on mixed-breed dogs is unusual but shrewd, says Adam Boyko, a geneticist at Cornell University in Ithaca, New York. Although more than half of the dogs in the United States are mongrels, genetic studies tend to focus on pure-bred animals. “Genetics often deals with the interactions between genes,” says Boyko. “And if you want to truly understand those, you want to study individuals where you’ve shuffled up the genes.”

Human participants in Darwin’s Dogs, which launched last October, answer about 130 questions about their pets’ behaviour. The questions cover everything from ‘Does your dog generally enjoy life?’ (the answer, says Karlsson, is overwhelmingly ‘yes’) to ‘Does your dog cross its paws when it lies down?’. Some questions were inspired by surveys that assess impulsivity in humans. Other questions have been suggested by Alonso, who is the executive director of the International Association of Animal Behavior Consultants in Cranberry Township, Pennsylvania, and by other dog trainers on the basis of observations made over decades of working with animals that have behavioural problems.

Karlsson says that she is thinking of expanding the list of questions even further.



Tail-chasing in dogs is suspected to share genetic roots with human obsessive-compulsive disorder.

“Fortunately, it turns out that people love to talk about their dogs,” she says.

Ultimately, the success of the project may hinge on the quality of those surveys and the specificity of the questions asked, says Wynne. Asking owners whether their dog is happy, for example, could yield mixed results. “One person’s unhappy dog is another person’s comfortably resting dog,” he says. “A good question would be: ‘Does your dog poop on the carpet?’ Because poop on the carpet is pretty damn clear.”

It is still unclear how useful the results from dogs will be in shedding light on human behavioural variation. Karlsson is hopeful that even if different genes are involved in the two species, they may converge on the same cellular pathways. Gerald Nestadt, a psychiatrist who specializes in OCD at Johns Hopkins University in Baltimore, Maryland, notes that affected animals often display only one type of compulsive behaviour, whereas a human with

OCD will typically have several.

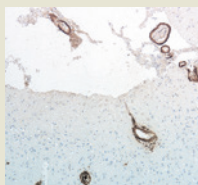
Even so, he adds, the field is hungry for any leads it can get. “Anything that will help is worth trying,” he says. “I think this project is a great idea.”

For their part, Alonso and other participants are eager to learn more about their own dogs and why they behave the way they do. Miranda Workman of Buffalo, New York, enrolled her three dogs — Zeus, Athena and Sherlock — into the study, in part to gain insight into their behavioural quirks. Although Athena, a 34-kilogram Dutch shepherd, was bred to be a dedicated herding and guarding dog, she has a jovial side that is not often found in her breed. And Sherlock, a Jack Russell, is more shy and sensitive than other terriers.

“I have some dogs that don’t necessarily fit the stereotype,” says Workman. “Is it their environment that’s different or are they different? It will be fun to find out why they are that way.” ■

MORE ONLINE

TOP STORY



More evidence for ‘transmissible Alzheimer’s’ theory
go.nature.com/jo5gnx

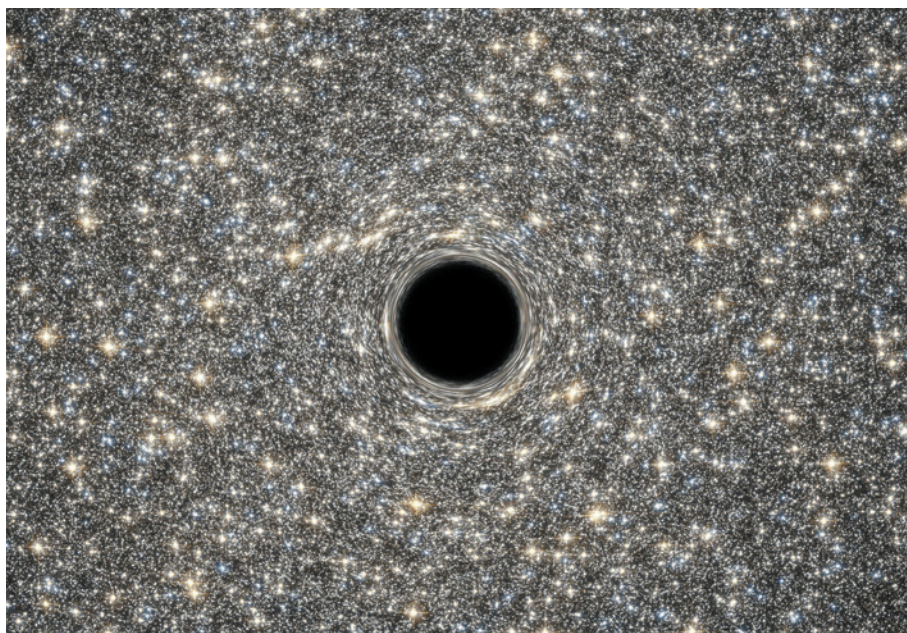
MORE NEWS

- Study tallies ‘true catch’ of global fishing go.nature.com/aallcz
- Researchers question design of fatal French clinical trial go.nature.com/phbt9v
- End in sight for decade-long chemistry paper controversy go.nature.com/6zm95u

NATURE PODCAST



The computer program that plays Go, a general ‘ageing’ factor and the stolen library of John Dee. nature.com/nature/podcast



A black hole, visualized here in the M60-UCD1 galaxy, was thought to lose information as it disappears.

PHYSICS

Physicists split by Hawking paper

Some welcome his latest report as a fresh way to solve a black-hole conundrum; others are unsure of its merits.

BY DAVIDE CASTELVECCHI, CAMBRIDGE, UK

Almost a month after Stephen Hawking and his colleagues posted a paper about black holes online¹, physicists still cannot agree on what it means.

Some support the preprint's claim — that it provides a promising way to tackle a conundrum known as the black hole information paradox, which Hawking identified more than 40 years ago. “I think there is a general sense of excitement that we have a new way of looking at things that may get us out of the logjam,” says Andrew Strominger, a physicist at Harvard University in Cambridge, Massachusetts, and a co-author of the latest paper.

Strominger presented the results on 18 January at a crowded talk at the University of Cambridge, UK, where Hawking is based.

Others are not so sure that the approach can solve the paradox, although some say that the work illuminates various problems in physics. In the mid-1970s, Hawking discovered that black holes are not truly black, and in fact emit some radiation². According to quantum physics, pairs of particles must appear out of

quantum fluctuations just outside the event horizon — the black hole's point of no return. Some of these particles escape the pull of the black hole but take a portion of its mass with them, causing the black hole to slowly shrink and eventually disappear.

In a paper³ published in 1976, Hawking pointed out that the outflowing particles — now known as Hawking radiation — would have completely random properties. As a result, once the black hole was gone, the information carried by anything that had previously fallen into the hole would be lost to the Universe. But this result clashes with laws of physics that say that information, like energy, is conserved, creating the paradox. “That paper was responsible for more sleepless nights among theoretical physicists than any paper in history,” Strominger said during his talk.

The mistake, Strominger explained, was to ignore the potential for the empty space to carry information. In their paper, he and Hawking, along with their third co-author Malcolm Perry, also at the University of Cambridge, turn to soft particles. These are low-energy versions of photons, hypothetical particles known

as gravitons and other particles. Until recently, these were mainly used to make calculations in particle physics. But the authors note that the vacuum in which a black hole sits need not be devoid of particles — only energy — and therefore that soft particles are present there in a zero-energy state.

It follows, they write, that anything falling into a black hole would leave an imprint on these particles. “If you're in one vacuum and you breathe on it — or do anything to it — you stir up a lot of soft gravitons,” said Strominger. After this disturbance, the vacuum around the black hole has changed, and the information has been preserved after all.

The paper goes on to suggest a mechanism for transferring that information to the black hole — which would have to happen for the paradox to be solved. The authors do this by calculating how to encode the data in a quantum description of the event horizon, known whimsically as ‘black hole hair’.

TRICKY TRANSFER

Still, the work is incomplete. Abhay Ashtekar, who studies gravitation at Pennsylvania State University in University Park, says that he finds the way that the authors transfer the information to the black hole — which they call ‘soft hair’ — unconvincing. And the authors acknowledge that they do not yet know how the information would subsequently transfer to the Hawking radiation, a further necessary step.

Steven Avery, a theoretical physicist at Brown University in Providence, Rhode Island, is sceptical that the approach will solve the paradox, but is excited by the way it broadens the significance of soft particles. He notes that Strominger has found that soft particles reveal subtle symmetries of the known forces of nature⁴, “some of which we knew and some of which are new”.

Other physicists are more optimistic about the method's prospects for solving the information paradox, including Sabine Hossenfelder of the Frankfurt Institute for Advanced Studies in Germany. She says that the results on soft hair, together with some of her own work, seem to settle a more-recent controversy over black holes, known as the firewall problem (see *Nature* **496**, 20–23; 2013). This is the question of whether the formation of Hawking radiation makes the event horizon a very hot place. That would contradict Albert Einstein's general theory of relativity, in which an observer falling through the horizon would see no sudden changes in the environment.

“If the vacuum has different states,” Hossenfelder says, “then you can transfer information into the radiation without having to put any kind of energy at the horizon. Consequently, there's no firewall.” ■

1. Hawking, S. W., Perry, M. J. & Strominger, A. Preprint at <http://arxiv.org/abs/1601.00921> (2016).
2. Hawking, S. W. *Nature* **248**, 30–31 (1974).
3. Hawking, S. W. *Phys. Rev. D* **14**, 2460–2473 (1976).
4. Strominger, A. *J. High Energy Phys.* **1407**, 152 (2014).

GENETICS

Monkeys genetically modified to show autism symptoms

But it is unclear how well the results match the condition in humans.

BY DAVID CYRANOSKI

The laboratory monkeys run obsessively in circles, largely ignore their peers and grunt anxiously when stared at. Engineered to have a gene that is related to autism spectrum disorder in people, the monkeys are the most realistic animal model of the condition yet, say their creators. Researchers hope that the animals will open up new ways to test treatments and investigate the biology of autism. But the jury is still out on how well the monkeys' condition matches human autism.

Autism has a vast array of symptoms and types, and researchers think that at least 100 genes play a part. The scientists who led the latest work, which is published on 25 January in *Nature* (Z. Liu *et al.* *Nature* <http://doi.org/bb3k>; 2016), turned to the autism-related gene *MECP2*: many of the symptoms of autism are found in people who have extra copies of the gene (*MECP2*-duplication syndrome) as well as in people who have certain mutations in this gene (Rett's syndrome). Researchers have engineered monkeys to have autism-related genes before (H. Liu *et al.* *Cell Stem Cell* **14**, 323–328; 2014), but this is the first published demonstration of a link between those genes and the animals' behaviour.

Back in 2010, the team that did the latest work, led by researchers at the Chinese Academy of Sciences' Institute of Neuroscience in Shanghai, attached human *MECP2* genes to a harmless virus, which they injected into the eggs of crab-eating macaque monkeys (*Macaca fascicularis*). The eggs were then fertilized, and the developing embryos were implanted into female monkeys. The result was 8 genetically manipulated newborns, which each had 1–7 extra copies of *MECP2*. Examinations of other, stillborn monkeys revealed that the extra copies were being expressed in the brain. "That was the first exciting moment," says Zilong Qiu, a molecular biologist at the Institute of Neuroscience and a co-author of the paper.

The next breakthrough came about a year later, when the monkeys showed behaviours that hinted at autism: running around in tight circles in a strange manner. "If another monkey is in its way, it will either jump over the monkey, or go around it, but then it would return to its original circular path," says co-author Sun Qiang, a reproductive biologist at the institute.

The team launched a battery of behavioural



A macaque made to have autism-like behaviours.

tests, which showed that all of the monkeys had at least one autism-like symptom, such as repetitive or asocial behaviour, and that the symptoms were more severe in males, as seen in people with the *MECP2* duplications. But this still wasn't enough to be sure that the monkeys were a sound model of autism — and a paper that the team submitted for publication in 2013 was rejected. Among other things, reviewers wanted to know whether the unusual behaviour was just the result of fiddling around with the genome. "We needed to show where the gene makes a difference," says Qiu.

That opportunity came with the next generation of macaques, which the team created with unprecedented speed. When the monkeys were 27 months old and not yet sexually mature, Sun's team took testes from the males, matured the tissue artificially by grafting it under the skin on the backs of castrated mice, and used the resulting sperm to fertilize eggs from non-engineered macaques. The offspring showed asocial behaviour at about 11 months. That both gene and symptoms seemed to be passed on to a second generation was finally enough to convince reviewers, says Qiu.

The macaque model is "superior" to mouse models of autism because "it actually shows more clearly some of the autism-like behaviours," says Alysso Muotri, who researches stem cells, autism and Rett's syndrome at the

University of California, San Diego. But he adds that the symptoms in both mice and monkeys still seem less severe than "what we actually observe in human patients". "It remains to be seen if the model can actually generate novel insights into the human condition," he says.

Huda Zoghbi, a pioneer of *MECP2* studies in mice at Baylor College of Medicine in Houston, Texas, is even more cautious. The monkeys do not mimic some of the human *MECP2*-duplication symptoms, such as seizures and severe cognitive problems, she notes. This could be because the expression of the gene in the monkey model is triggered by a different mechanism from that in humans — a limitation that the authors recognize — and she advises caution in using the model to make assumptions about human autism.

Qiu, meanwhile, is excited by the prospect of using the model to identify exactly where in the brain the *MECP2* overexpression causes trouble. His team is already using brain-imaging technology on the monkeys to pinpoint such areas. Next, the researchers plan to use the CRISPR gene-editing technique to knock out the extra *MECP2* copies in cells in those regions and then check whether the autism-like symptoms stop.

It is unlikely that such a technique would be approved for use in people any time soon. But the regions identified in the monkey study could be targeted with other, existing treatments — such as deep brain stimulation, which has

had success in treating Parkinson's disease and depression. Because the structure of the mouse brain is so different from that

of the human brain, Qiu says that the monkey imaging will allow more parallels to be drawn with humans than mice studies could. Working with a mental-health hospital, the team is also trying to identify the autism-linked genes that are most common in the Chinese population.

If non-human primates prove to be a useful model for psychiatric disorders, China and other countries that are investing heavily in research on monkeys, such as Japan, could gain an edge in brain research. Muotri says that such studies probably wouldn't be done in the United States, where research on monkeys is more expensive and controversial. "China and Japan have a clear advantage over the US on this area," he says. ■

"We needed to show where the gene makes a difference."

CLIMATE

2015 breaks heat record

Pacific Ocean warming helped to make last year the hottest in history.

BY JEFF TOLLEFSON

It's official: 2015 was the hottest year on record. Global data show that a powerful El Niño system, marked by warmed waters in the tropical Pacific Ocean, helped to drive atmospheric temperatures well past 2014's record highs. Some researchers suggest that broader Pacific trends could spell even more dramatic temperature rises in years to come.

Released on 20 January, the temperature data come from independent records maintained by NASA, the US National Oceanic and Atmospheric Administration and the UK Met Office. All three, along with an analysis released by the World Meteorological Organization on 25 January, document unprecedented high temperatures in 2015, pushing the global average to at least 1 °C above pre-industrial levels and 0.16 °C above 2014.

Although El Niño contributed to the warmth late in the year, US government scientists say that the steady increase in atmospheric concentrations of greenhouse gases continues to drive overall warming.

The current El Niño is predicted to continue to boost average global temperatures over the next several months. This could translate into another year of record heat. But the question facing scientists is whether this near-record El Niño has helped to flip the Pacific Ocean into a warmer state.

The Pacific Decadal Oscillation (PDO) is a 15- to 30-year cycle that increases sea surface temperatures across the eastern Pacific in its positive phase and produces cooler temperatures in its negative phase. Since 1998, after the last major El Niño and a subsequent La Niña cooling, the PDO has been mostly negative. Some scientists say that the cooling helped to suppress the increase in global temperatures in the early part of the millennium. But since early 2014, the PDO has been mostly positive.

Jerry Meehl, a climate modeller at the National Center for Atmospheric Research in Boulder, Colorado, has a study under review that suggests that the PDO is likely to remain in a positive state over the coming decade. "Over the next ten years," he says, "we see higher rates of warming." ■



Fires in Indonesia last year contributed greatly to the nation's greenhouse-gas emissions.

CLIMATE

Paris deal strains carbon accountancy

Developing countries need help to meet reporting rules.

BY JEFF TOLLEFSON

Indonesia's Central Kalimantan province, a lightly populated swathe of Borneo, is a hotbed for greenhouse gases that are emitted through deforestation, family farming and industrial palm-oil production. Last year, these activities fuelled devastating fires that torched more than 400,000 hectares in the province and at least 2.5 million hectares across the developing nation. By some estimates, the fires released into the atmosphere more than twice as much carbon as Germany emits in a year.

Calculating the volume of greenhouse gases emitted across such a dynamic landscape is not an easy task. Nevertheless, the climate agreement made in Paris last month dictates that nearly every country will need to begin assembling detailed inventories of their greenhouse-gas emissions in a few years' time. In Indonesia, the national government has delegated much of that responsibility to provincial governments. Soon, administrations that often struggle to provide basic public services will be required to master the complex science of carbon accountancy.

"We cannot rely on the local governments," says Rizaldi Boer, who heads the Centre for Climate Risk and Opportunity Management in Southeast Asia and the Pacific in West Java. "We need to

integrate this kind of training into the local universities."

It is a challenge that is being faced by developing countries across the globe. The Paris agreement relies on a 'pledge and review' programme to reduce emissions and halt global warming over the course of the twenty-first century. Under that strategy, countries must document their progress towards voluntary commitments to limit carbon emissions. Solid and transparent data will be needed to verify that they are living up to their promises.

"This is really the compliance regime for the Paris agreement," says Alden Meyer, director of strategy and policy for the Union of Concerned Scientists in Washington DC. "This is how you really tell how well you are doing."

Developed countries have been submitting detailed reports on greenhouse-gas emissions to the United Nations for years. But until now, developing countries — which produce nearly two-thirds of global greenhouse-gas emissions — were not required to provide such comprehensive reports on a regular basis. Under the Paris agreement, most countries will need to supply inventories of greenhouse-gas emissions every two years — and although the deal includes some flexibility, the details have yet to be resolved.

Efforts are under way to build a network of professional carbon accountants across

NOVA WAHYUDI/XINHUA/CORBIS

the developing world. Partnering with the Greenhouse Gas Management Institute (GHGMI), a non-profit organization headquartered in Washington DC that has trained more than 3,000 people in carbon accounting, Boer is developing a formal curriculum for universities in Indonesia. His focus lies on teaching the assessment of land-use emissions, which are the largest — and hardest to quantify — source of greenhouse gases in the country. The ultimate aim is to build a qualified workforce to feed into local governments as well as businesses and other institutions that are working to tackle climate change.

“We can’t implement this agreement without building capacity,” says John Niles, who directs the Carbon Institute, an arm of the GHGMI that develops training programmes in the measurement and monitoring of carbon emissions. “We need the right investments in the right institutions, in every country on Earth.”

LOCAL RESPONSIBILITY

Although the central government of Indonesia provides basic data on deforestation, provincial officials must calculate emissions from the expanding agriculture sector as well as from the energy, industrial and municipal sectors. After the data have been collected and the calculations performed, Boer says,

officials must complete and submit 186 forms to meet the reporting requirements laid out by the Intergovernmental Panel on Climate Change.

“It’s very hard for the local governments to understand,” he says. But if Indonesia

“We can’t implement this agreement without building capacity.”

succeeds in building the technical capacity to conduct such assessments, it will become easier for the country to test new policies and to assess what will work

best to reduce emissions in the future. “We have to see this as an opportunity,” says Boer.

Cost will be a barrier. Although poorer countries such as Indonesia finally agreed to a unified framework for reporting emissions in Paris, they fought hard for assurances that wealthier countries would provide money to help them to kick-start the process.

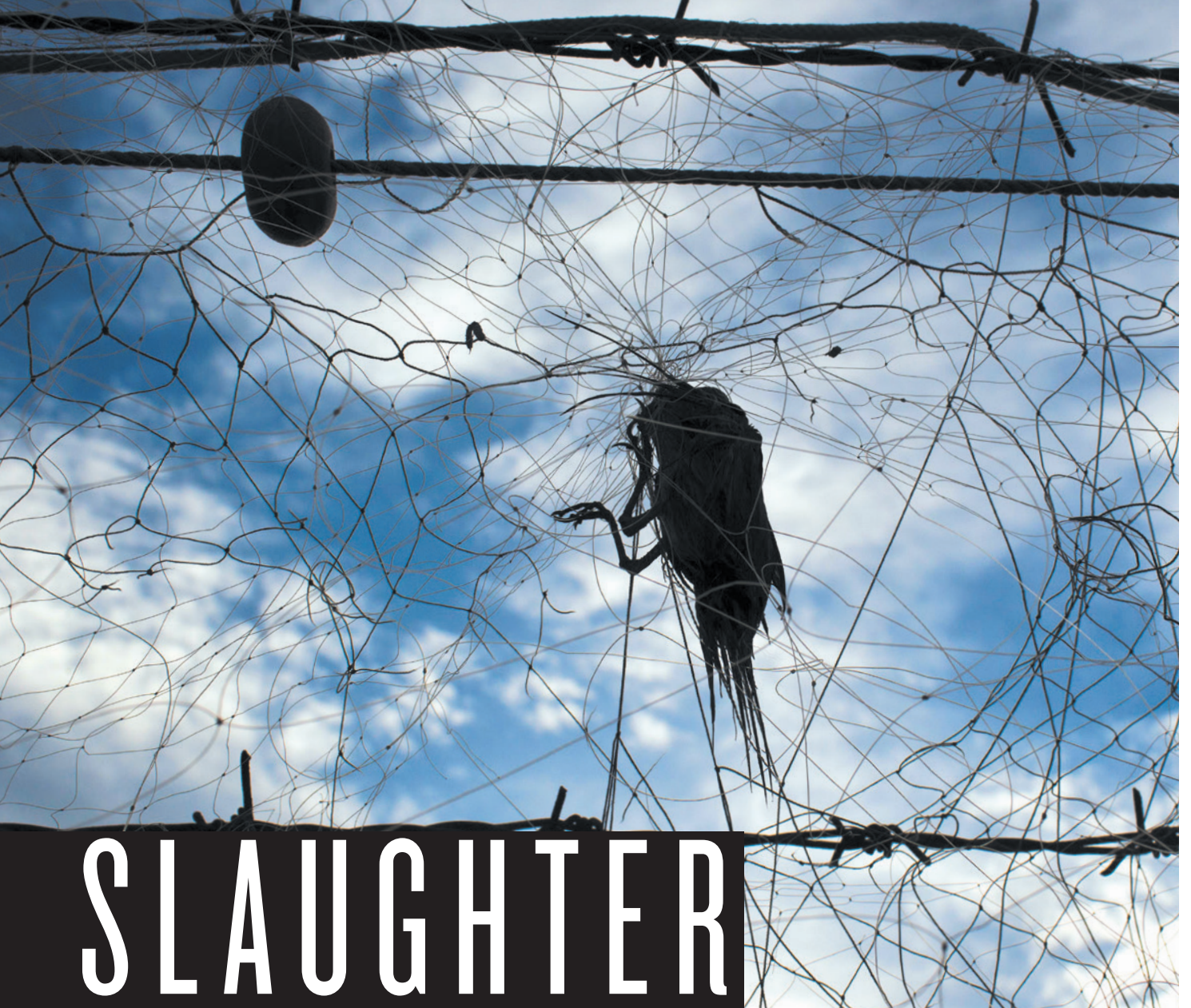
Various initiatives are already under way. The GHGMI has developed an online training course in carbon accounting and is investigating ways to provide ongoing support for individuals and institutions in developing countries. And in March, a coalition of countries and climate-advocacy organizations led by Germany is expected to launch a US\$15-million initiative to help less-wealthy

countries prepare to monitor and report their greenhouse-gas emissions.

The process could function differently in each country. But Yamide Dagnet, a former climate negotiator for the UK government who works for the World Resources Institute in Washington DC, suggests that academic institutions can serve as repositories for expertise on behalf of governments, which often lose their most-skilled and experienced carbon accountants to the United Nations, think tanks and corporations. “You could have universities and institutes designed to provide data, with some permanent experts and graduate students as well,” says Dagnet. “I think we need to be really creative.”

Perhaps the biggest challenge is how to move quickly. Developing countries may need to begin reporting within a few years, says Michael Gillenwater, executive director and dean of the GHGMI. The prevailing approach of hosting workshops and one-time training sessions to promote minimal local expertise while bringing in consultants to oversee one-off greenhouse-gas inventories will not be enough.

“The current model will break if you try and scale it up,” says Gillenwater. “We need more and better people and we need a different model.” ■



DAVID GUTTENFELDER/AP/NATIONAL GEOGRAPHIC CREATIVE

SLAUGHTER of the songbirds

Mist nets are strung across flight paths to trap birds.

Songbirds are a culinary delicacy in Cyprus — but catching and eating them is illegal. Even so, the practice is on the rise and could be threatening rare species.

BY SHAONI BHATTACHARYA

It wasn't until I saw the blade glinting in the sunlight that I realized how grave the situation was. Broad and belligerent in army fatigues, the man strode along the track, ranting in Greek. Behind his back, his hands flexed a knife blade in and out of its wooden handle. This man was a trapper, a poacher of birds — and he clearly didn't want company. "What are you doing here?" he demanded.

My companions and I had come to this dry scrubland on the Mediterranean island of Cyprus to look for evidence of songbird trapping. The birds are caught illegally and eaten in a traditional dish called *ambelopoulia* — and I was joining a September trip to monitor the extent of trapping. With me was Roger Little, a British conservation volunteer, and Savvas, a field officer with the conservation group BirdLife Cyprus whose name has been changed to protect his identity. We didn't expect to encounter trappers at this spot in the southeastern region of Cape Pyla; they usually work at night, when the birds are active. But now it seemed that they had started patrolling the site during the day. "You are on my land," the trapper said to us in Greek.

"If this is your property, then I apologize — we didn't know, we are going," Savvas said. We acted casual as the man escorted us back to the

SOURCE: BIRDLIFE CYPRUS

battered four-by-four in which we had come. “I shouldn’t really be letting you go,” he muttered. Moments later, we were driving away.

Bird trapping in Cyprus has grown into a controversy that encompasses crime, culture, politics and science. The practice was made illegal more than 40 years ago — but that simply forced it underground. Today, trappers routinely cut wide corridors through vegetation and string fine ‘mist nets’ from poles to catch the birds, which are sent to local restaurants and quietly served. A platter of a dozen birds sells for €40–80 (US\$44–87), and the trade in songbirds is responsible for an estimated annual market of €15 million. The delicacy is so prized and lucrative that it is suspected to be linked to organized crime, and those trying to stop it have been subject to intimidation and violence.

Conservation organizations say that the trapping is increasing and that it is threatening rare bird species that stop in Cyprus during their migration. Last March, a report by BirdLife Cyprus suggested that some 2 million birds had been killed in the previous autumn, including 78 threatened species. The group claims that trapping — on top of threats from climate change, habitat loss and invasive species — could cause irreparable damage to some bird populations. “Illegal bird killing just cannot be justified, it’s like the last kick off the cliff for some species,” says Claire Papazoglou, executive director of BirdLife Cyprus near Nicosia.

But the picture is not black and white, in part because the extent of bird killing is disputed and its effects on bird populations are unclear. Critics have questioned the methods used by BirdLife Cyprus to estimate the numbers being captured on the island. The debate led to a workshop last July to discuss the science, with representatives from all agencies involved. Attendee Alison Johnston, an ecological statistician at the British Trust for Ornithology (BTO), a charitable research institute in Thetford, says that so little is known about the population sizes and routes of migratory birds in the Mediterranean that it is difficult to assess the full impacts of trapping. “If we knew more about the numbers,” she says, “we could say whether this is a critical number being killed.”

The debate over Cyprus’s songbirds could have wider repercussions, because bird killing is rife in other parts of the world. A 2015 report from BirdLife International estimates that hunters are killing about 25 million birds a year over the whole Mediterranean region; Cyprus stands out because so many are killed in such a small country. Globally, more than half of the world’s migratory bird populations are thought to be in decline. “This isn’t just an issue for Cyprus, or Africa, or Europe,” says Claire Runge, a conservation scientist at the University of Queensland in Brisbane, Australia, who led a study published last December showing that only 9% of migratory birds worldwide are adequately protected across their range. “Countries will need to work together to find a solution to what is essentially a human–wildlife conflict,” she says.

Papazoglou worries that what is happening in Cyprus sets a dangerous precedent. “This level of rampant illegality in an EU country sends a terrible message to the rest of the world. If rich, stable and well-run countries cannot enforce wildlife law, what hope is there to get fragile countries in the Middle East and Africa to act?”

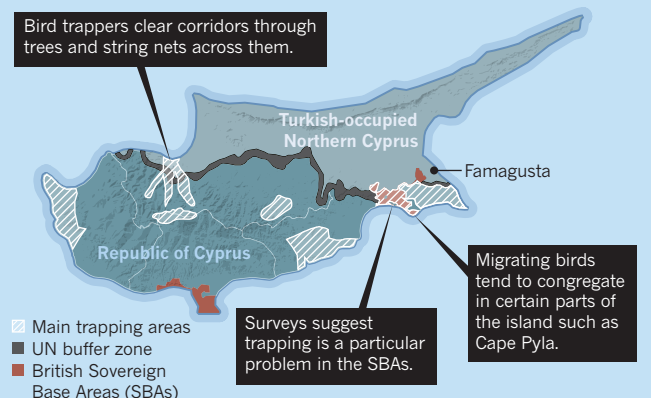
GATEWAY TO CONTINENTS

Situated in the far southeastern corner of the Mediterranean, Cyprus is a gateway to three continents and has been fought over for millennia (see ‘Trapped in Cyprus’). It is currently sliced up into four jurisdictions: the Republic of Cyprus and the Turkish-occupied region of Northern Cyprus, separated by a UN buffer zone, and two small pockets called Sovereign Base Areas (SBAs) that were retained by the United Kingdom after the island gained independence in 1960 because of their strategic military importance. (Britain is currently using one of these areas to deploy air strikes to Syria.)

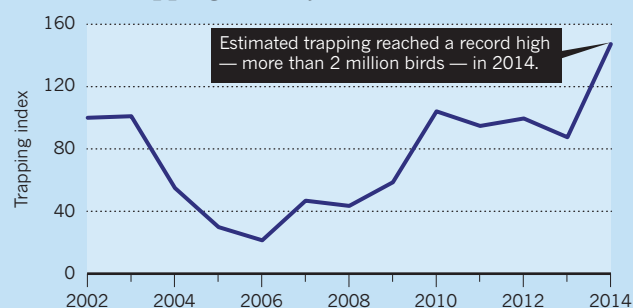
The island’s location also makes it an ideal rest-stop for migratory birds. Nearly half of the bird species from Europe, North Africa and the Middle East are thought to use the island as a migratory staging post as they fly south in the autumn, and back again in spring. These include common birds such as sparrows and the European robin (*Erithacus rubecula*), as well as threatened species including the barn owl (*Tyto alba*), common

TRAPPED IN CYPRUS

Cyprus is a key stop for migratory birds and a hotspot for illegal bird trapping. Surveys estimate that trapping is on the rise.



Autumn trapping activity



kingfisher (*Alcedo atthis*) and European turtle dove (*Streptopelia turtur*). All of these creatures have been found in the trappers’ nets, as have some threatened, non-migratory, endemic bird species such as the Cyprus warbler (*Sylvia melanothorax*) and the Cyprus wheatear (*Oenanthe cypriaca*).

The practice of trapping dates back to a time when birds were among the few easily found sources of protein on this arid island. Originally, *ambelopoulia* would have been a plate of blackcaps (*Sylvia atricapilla*), but the dish has extended to include 22 species of songbird. The traditional trapping method is to ensnare birds in trees with strategically placed ‘limesticks’ — twigs coated in a goo of mud mixed with Syrian plum juice. But in 1974, laws were introduced to ban non-selective capture methods, including limesticks and mist nets. Bird trapping is also illegal under the European Union (EU) Birds Directive and the Convention on the Conservation of European Wildlife and Natural Habitats (known as the Bern Convention), both of which Cyprus has adopted.

The practice never stopped. Many Cypriots argue that bird trapping for *ambelopoulia* is a tradition and a right, and it has become a highly emotive issue. In the Famagusta district, raids on restaurants and arrests related to bird trapping have sparked public protests, and some politicians either covertly or overtly support it. Last December, Evgenios Hamboulas, Famagusta member of parliament for the incumbent Democratic Rally party, posted a photo of himself on Facebook seated in front of a plate of songbirds with the caption: “Soon in our restaurants! Happy holidays!” The post received nearly 600 likes in 5 days, and condemnation from his party.

Conservation groups believe that bird trapping is rising fast; last year’s BirdLife Cyprus report said that the practice had reached “industrial scale”. Trappers rip out the island’s native scrub bushes, then plant and irrigate lush, bright-green acacia trees that attract birds. They cut corridors through the groves and string mist nets across them from poles.

DNA identifies baked birds

A DNA technique used to identify the provenance of foods is being turned into the latest crime-fighting tool to tackle illegal bird killing in Cyprus.

About 22 bird species are served up illegally by Cypriot restaurants in a traditional cuisine called *ambelopoulia* (pictured). The hope is that DNA barcoding — a technique that uses DNA sequences to identify a species — can show whether a restaurant owner is passing off illegally trapped birds as chicken or another meat during a raid. Proof that the birds being eaten are protected species might help to build a stronger case to prosecute lawbreakers.

The project is a three-year collaboration between researchers at the University of Pisa, Italy, and the University of Cyprus, Nicosia, as well as the Ministry of the Interior's Cyprus Game and Fauna Service and the conservation group BirdLife Cyprus. So far, the team's unpublished work has shown that sequences from part of a single gene (for cytochrome c oxidase) are enough to distinguish 81 bird species. This worked even when the DNA was extracted from meat baked at 90°C, and cooked with salt or vinegar — a method that matches local gastronomy but that could degrade the DNA. "[It was] prepared in a particular way so we would be sure our DNA investigations would be effective," says Filippo Barbanera, a zoologist at the University of Pisa.

The Pisa team has helped to set up a molecular-genetics lab at the University of Cyprus to do the DNA analyses. It has also trained two game-service officers, so that they can testify about the forensic DNA evidence in court. So far, the method has been used in two cases that are pending court, says Panicos Panayides, an officer at the game service in Nicosia. "We have a new card to be played to stop, or to try to reduce, the rate of illegal trade and consumption of birds in Cyprus," says Barbanera. **S.B.**



When Savvas, Roger and I stopped at a well-known trapping hotspot, the evidence was everywhere: metal poles were concreted into bases made of empty tyres; black irrigation pipes criss-crossed the dusty earth; old carpets covered the ground to stop vegetation growing where the nets hang.

Earlier in the trip, we found an MP3 player high in an acacia tree broadcasting a repetitive birdsong — a 'tape lure' used to attract the birds. Nearby, a red-backed shrike (*Lanius collurio*) and a sparrow both flailed frantically, their feet and wing tips glued onto limesticks balanced high in the tree.

DEATH TOLL

Conservationists first started systematically monitoring the extent of bird trapping in 2002, using a protocol developed by BirdLife Cyprus and the United Kingdom's Royal Society for the Protection of Birds,

in consultation with the Cyprus Game and Fauna Service (part of the Ministry of Interior) and the British SBA police. The figures showed an initial dip in trapping around the time that Cyprus acceded to the EU, then an upward trend from 2007. But the 2014 trapping figures, published last year, caused a particular stir.

The 2 million birds that BirdLife Cyprus estimated were captured during the previous trapping season was the biggest jump since monitoring had begun. The report also broke down trapping trends by jurisdiction, and found that the SBAs accounted for much of the increase. It estimated that 900,000 birds were killed there — even though the regions take up only 3% of the land — and that there had been a 199% increase since 2002. By contrast, the Republic of Cyprus had seen a downturn in illegal trapping. (Bird killing is not thought to be a major problem in Northern Cyprus.)

The record-breaking numbers prompted criticism and headlines, and led some conservationists and media to imply that the British authorities were turning a blind eye to trapping so as not to upset the local community. The SBA Administration told BirdLife Cyprus that it "does not accept the survey findings" and questioned some of the figures in the report. According to people at the July meeting, the administration was particularly concerned with how the estimates were reached.

The main trapping survey is carried out over a six-week period in the autumn migratory season — also the main bird-hunting season. The surveillance team regularly visits 60 sites, each one kilometre square, that are deemed prime trapping territory and assigns them one of five categories on the basis of the scale of mist netting that it observes — from 'active set net' (where trappers have left a net unfurled on poles), to 'prepared' (where undergrowth has been freshly cut to produce a corridor, but no nets are present), to 'clear' (areas with no evidence of trapping).

From these data, the team estimates how many birds are killed in the region and season overall. To do this, it must make assumptions, such as the number of birds caught in a net each day and that bird migration is relatively constant, when in reality it occurs in waves. "We always say that our estimate of numbers caught is full of assumptions," Papazoglou says. "It needs to be read with a lot of caveats."

One of the major contentions of the SBAs is over the 'prepared' category, Johnston says — because deciding whether an area is about to be used for trapping is to some extent subjective. And others have expressed concerns about the accuracy of the estimates. "We have some doubts over the specifics of the monitoring and the exact numbers," says Panicos Panayides, an officer at the Cyprus Game and Fauna Service in Nicosia.

The July workshop was convened to address these methodology issues. BirdLife Cyprus invited Johnston, her colleague Nick Moran, who runs a major British bird survey, other bird-monitoring experts and representatives of the SBAs. After the workshop, Johnston and Moran advised BirdLife Cyprus to do away with the 'prepared' category and to increase the number of squares sampled within the SBAs, among other recommendations. BirdLife Cyprus will adopt these in its 2016 analysis, which should be published this spring. In a statement to *Nature*, the SBA Administration said that it did not wish to comment on the methodology used previously, that "all groups are working together to refine the recommendations produced by the BTO", and that "it is imperative that we continue to work together to counter the practice [of bird trapping]".

But even with some adjustments, the trapping figures still jumped between 2013 and 2014, says Johnston. "The [new] equation slightly reduces the estimated number, but not by much, by about 10%." And the year-on-year trend towards increased trapping is sound, she says, because the monitoring methods have been consistent over time. If anything, she thinks that BirdLife Cyprus's estimated numbers are "conservative". On the basis of previous studies, the group estimates that about 20 birds are captured in each net per day. But this figure could be much higher if, as is common today, trappers use

taped songs to attract birds. One study² estimated that such lures can increase the number of birds flying into traps by up to 13-fold.

Elsewhere in the world, hunting is thought to be playing a part in the demise of even common bird species. Last year, researchers warned that a highly abundant Eurasian bird, the yellow-breasted bunting (*Emberiza aureola*), had lost as much as 95% of its population in the past three decades or so and was close to extinction in parts of its range. One major driver is thought to be the trapping of birds in China, where they are served as an expensive delicacy³.

Accurately measuring the extent of bird killing is important if researchers and conservationists are to gauge the damage being done to bird populations, and to encourage efforts to clamp down. But Johnston says that getting rock-solid data is extremely difficult — particularly when visiting the monitoring sites is fraught with danger. “If the trappers had to fill in a form and say how many birds they caught on different days — we could do a great analysis,” she says.

Runge says that low-level hunting of common species may not have a huge impact on populations. “For other endangered species, where only a few individuals are left, it can be really critical.” And whatever the precise numbers, all the agencies involved agree that bird killing in Cyprus needs to be tackled. The question is, how.

POLITICAL SENSITIVITIES

Jim Guy, divisional commander of the eastern SBA police, is polite, charming and hard as nails. I met him at the police station in Dhekelia, a cluster of low-lying buildings behind a wire fence set off the road, a few kilometres from the city of Larnaka. He’d originally come from Glasgow on a 3-year posting, but has ended up staying for 17.

“As far as the bases themselves are concerned, there’s no denying it’s one of the main trapping areas,” he says. But Guy seems aggrieved about the criticism aimed at the SBAs since BirdLife Cyprus’s report, and says that lax enforcement is not to blame. Rather, he says, the eastern SBA — and especially the promontory of Cape Pyla — is a target for trappers because it is a key stopping point in the flight path of migratory birds. “Cape Pyla in particular has no buildings or houses or anything to deter, or put off birds, so it’s an ideal situation.”

Guy says that his team takes a three-pronged approach to tackle trappers: prevention, education and enforcement. “To some extent, enforcement is an Elastoplast,” he says. It might catch some trappers, but the practice will continue as long as there is demand for high-priced *ambelopoulia* from diners and the restaurants that serve it — and these lie almost entirely in the Republic.

Stopping that demand is extremely difficult, Guy adds. “The illegal practice in some cases is overtly or very often tacitly supported by people in very high political and administrative positions.” What’s more, officers trying to tackle trapping can find themselves threatened or worse. “In the UK, you can go home at night and you don’t have to think about your home or your family being attacked,” says Guy, who has had officers seriously assaulted while dealing with trappers.

His sense of frustration is shared by Panayides. The walls of his office are lined with pictures of birds, and an EU Birds Directive poster perches above the table. Panayides says that there have been at least 30 cases in the past decade in which game-service officers responsible for wildlife enforcement in the republic were harassed by trappers. “We’ve had people put bombs in the private cars of game wardens, and cases where the houses of game wardens have been burnt down,” he says.

Even when trappers are caught, Panayides says, the weak punishments imposed by courts are not effective deterrents. Technically, Cypriot law allows a first-time trapper to be jailed for up to 3 years, or fined up to €17,000. In reality, most get off with a fine of a few hundred euros. Panayides tells of one poacher whom his team has caught and prosecuted eight times over the past decade. “What else can we do as a department?” he says disconsolately.

The fight escalated last year. In May, a previously agreed plan to deal with bird killing was passing through Cyprus’s Council of



Blackcaps (top) and European bee eaters are both trapped in nets in Cyprus.

Ministers when the government added a last-minute clause that would allow selective hunting of blackcaps for *ambelopoulia*. The move caused an outcry in environmental organizations, because any method used to capture blackcaps would inevitably catch other species and is in breach of the Birds Directive. In August, the altered plan was rejected by the European Commission in a letter to the Cyprus government, and observers are now waiting to see how the government will respond.

Meanwhile, authorities in both the republic and the SBAs are stepping up efforts to curb bird killing. The republic authorities are looking at the use of a genetic technique known as DNA barcoding to identify the birds served up at restaurants (see ‘DNA identifies baked birds’), and the SBA Administration says that it removed 11 football-pitches’ worth of planted acacia from the central poaching area of Cape Pyla last summer. The removal met with demonstrations, and people sat in the dirt tracks to stop the clearance contractors. In the area where we encountered the knife-bearing poacher, the monitoring team now enters only if it has a police escort. The conservationists and the poachers have reached “the top-end of the fight”, says Savvas, who has been monitoring trapping on the island for nearly five years.

Surveillance and enforcement will only go so far: most parties agree that the only real way to tackle bird killing is through education and social change. “The general public has to recognize that this is not correct,” says Panayides. “Not just legally, but also morally and socially.” Papazoglou, too, is realistic about what needs to be done. “If we don’t get the minds and hearts of people to change — we will never change it,” she says. ■

Shaoni Bhattacharya is a science writer in London.

1. Runge, C. A. *et al. Science* **350**, 1255–1258 (2015).
2. Schaub, M. *et al. Auk* **116**, 1047–1053 (1999).
3. Kamp, J. *et al. Conserv. Biol.* **29**, 1684–1694 (2015).



There are at least six things in this picture that a quality-assurance manager would try to improve. Can you spot them?

QUALITY TIME

IT MAY NOT BE SEXY, BUT QUALITY ASSURANCE IS BECOMING A CRUCIAL PART OF LAB LIFE.

BY MONYA BAKER

R

ebecca Davies remembers a time when quality assurance terrified her. In 2007, she had been asked to lead accreditation efforts at the University of Minnesota's Veterinary Diagnostic Laboratory in Saint Paul. The lab needed to ensure that the tens of thousands of tests it conducts to monitor disease in pets, poultry, livestock and wildlife were watertight. "It was a huge task. I felt sick to my stomach," recalls Davies, an endocrinologist at the university's College of Veterinary Medicine.

She nevertheless accepted the challenge, and soon found herself hooked on finding — and fixing — problems in the research process. She and her team tracked recurring tissue-contamination issues to how containers were being filled and stored; they traced an assay's erratic performance to whether technicians let an enzyme warm to room temperature; and they established systems to eliminate spotty data collection, malfunctioning equipment and neglected controls. Her efforts were crucial to keeping the diagnostic lab in business, but they also forced her to realize how much researchers' work could improve. "That is the beauty of quality assurance," Davies says. "That is what we were missing out on as scientists."

Davies wanted to spread the word. In 2009, she got permission and financial support to launch an internal consulting group for the college, to help labs with the dry but essential work of quality assurance (QA). The group, called Quality Central, now supports more than half a dozen research labs — helping them to design systems to ensure that their equipment, materials and data are up to scratch, and helping them to improve.

She is also part of a small but growing group of professionals around the world who hope to transform basic biomedical research. Many were hired by their universities to help labs to meet certain regulatory standards, but these QA consultants have a broader vision. They are not pushing for universal adoption of formal regulatory certifications. Instead, they advocate 'voluntary QA'. With the right strategies, they argue, scientists can strengthen their research and improve reproducibility.

When Davies first started proselytizing to her fellow faculty members, the responses were not encouraging. "None of them found the idea compelling at all," Davies recalls. How important could QA be, they asked, if the US National Institutes of Health did not require it? How could anyone afford to spend money or time on non-essentials? Shouldn't they focus on the discoveries lurking in their data, and not the systems for collecting them?

But some saw the potential, based on their own experiences. Before she had heard of Quality Central, University of Minnesota virologist Montserrat Torremorell was grateful when a colleague let her use his instruments to track transmissible disease in swine. But the results made no sense. Samples from pigs experimentally infected with influenza showed extremely low levels of the virus. It turned out that her benefactor had, like many scientists, skimped on equipment maintenance to save money. "It was a real eye-opener," Torremorell recalls. "It just made me think that I could not rely on other people's equipment."

QUALITY FOR ALL

Quality systems are an integral part of most commercial goods and services, used in manufacturing everything from planes to paint. Some labs that focus on clinical applications implement certified QA systems such as Good Clinical Practice, Good Manufacturing Practice and Good Laboratory Practice for data submitted to regulatory bodies. There have also been efforts to guide research practices outside these schemes. In 2001, the World Health Organization published guidelines for QA in basic research. And in 2006, the British Association of Research Quality Assurance (now simply the RQA) in Ipswich issued guidelines for basic biomedical research. But few academic researchers know that these standards exist (Davies certainly didn't back in 2007).

Instead, QA tends to be ad hoc in academic settings. Many scientists are taught how to keep lab notebooks by their mentors, supplemented perhaps by a perfunctory training course. Investigators often improvise ways to safeguard data, maintain equipment or catalogue and care for experimental materials. Too often, data quality is as likely to be assumed as assured.

Scientific rigour has taken a drubbing in the past few years, with reports that fewer than one-third of biomedical papers can be reproduced (see *Nature* <http://doi.org/477>; 2015). Scientific culture, training and incentives have all been blamed for promoting sloppy work; a common refrain is that the status quo values publication counts over careful experimentation and documentation. "There is chaos in academia," says Masha Fridkis-Hareli, head of ATR, a biotechnology consultancy in Worcester, Massachusetts, that also conducts laboratory work to help move basic research into industry. For every careful researcher she has encountered, there have been others who have thought nothing of scribbling data on paper towels, repeating experiments without running controls and guessing at details months after an experiment. Davies insists that plenty of scientists are doing robust work, but there is always room for improvement (see 'Solutions'). "There are easy fixes to situations that shouldn't be happening, but are," she says.

➔ NATURE.COM
How does your lab's QA measure up?
go.nature.com/e6xupg

Michael Murtaugh, a swine biologist at the University of Minnesota, had tried to establish practices to beef up the reliability of his team's lab notebooks, but the attempts that he made on his own never gained traction. Then Davies got on his case. After a year or so of her "planting seeds" — as she puts it — Murtaugh agreed to work with Quality Central and implement a low-tech but effective solution.

On designated Mondays, each member of Murtaugh's lab draws a name from a paper bag to determine whose notebook to audit. The scientists check that their assigned books include relevant controls for experiments, and indicate where data are stored and which particular machine generated them. The group also makes sure that any problems noted in the previous check have been addressed. It takes about

ten minutes per researcher every few weeks, but that's enough to change people's habits. Graduate student Michael Rahe says that the checks ensure that he keeps his notebook legible and up to date. "I never used to put in raw data," he says.

Albert Cirera, a technologist developing gas nanosensors at the University of Barcelona in Spain, has also embraced QA. As his lab group grew to 12 people, he found it difficult to monitor everyone's experiments, and his own efforts to implement a tracking system were inadequate.

He turned to a university-based QA consulting service for help. Now, samples, equipment and their data are all linked with tracking numbers printed on stickers and recorded in individuals' notebooks, on samples and in a central tracking file. The system does not slow down experiments, and staying abreast of projects is a breeze, says Cirera. But getting to this point took about four months and frequent consultations. "It was not something that you can create from zero," he says.

MAKING A MARKET

Any scientist adopting a QA system has to wager that the up-front hassle will pay off in the future. "It is very difficult to get people to check and annotate everything, because they think it is nonsense," says Carmen Navarro-Aragay, head of the University of Barcelona quality team that worked with Cirera. "They realize the value only when they get results that they do not understand and find that the answer is lurking somewhere in their notebooks."

Even when experiments go as expected, quality systems can save time, says Murtaugh. Methods and data sections in papers practically write themselves, with no time wasted in frenzied hunting for missing information. There are fewer questions about how experiments were done and where data are stored, says Murtaugh. "It allows us to concentrate on biological explanations for results."

The more difficult data are to collect, the more important a good QA system becomes. Catherine Bens, a QA manager at Colorado State University in Fort Collins, says that she remembers getting cold, wet and dirty when she had to monitor a study involving ultrasound scans and blood samples from a population of feral horses in North Dakota. Typical animal-identification practices such as ear tagging were not allowed. So, before the collection started, Bens supported researchers as they rehearsed procedures, pre-labelled tubes, made back-up labels and recruited animal photographers and park volunteers to ensure that samples would be linked to the correct animals. Even in a snow storm with winds so loud that everyone had to shout, the team made sure that each data point could be traced.

Rare samples or not, few basic researchers are clamouring to get QA systems in place. Most are unfamiliar with the discipline, says Davies. Others are hostile. "They see it as trying to constrain them, and that you're making them do more work."

Before awarding certain grants, the Found Animals Foundation in

"THERE ARE EASY FIXES TO SITUATIONS THAT SHOULDN'T BE HAPPENING, BUT ARE."

Los Angeles, California, which funds research on animal sterilization, requires proof that instruments have been calibrated and that written plans exist for tracing data and dealing with outliers. It can be a struggle, says Shirley Johnston, scientific director of the foundation. One grant recipient argued that QA systems were unnecessary because just looking over the data would reveal their quality.

Part of the resistance may be down to how some QA professionals present themselves. “A lot of them are there to tell you what you are doing is wrong, and a lot of them are not very nice about it,” says Terry Nett, a reproductive biologist at Colorado State University who experienced this first-hand when he worked with outside consultants to incorporate Good Laboratory Practice principles in his lab. The effort was frustrating. “Instead of helping us understand, they would act like a dictator,” Nett recalls. “I just didn’t want them in my lab.” A few years ago, however, the university hired its own quality managers, and things changed. The current manager, Bens, acts more like a partner, Nett says. She points out where labs are already using robust practices, and explains the reasoning behind QA practices that she introduces.

To win scientists over, Bens stresses that QA systems produce data that can withstand criticism. “You build a support system around any data point you collect,” she says. When there is a strange result, researchers have documentation to trace its provenance. That can show whether a data point is real, an outlier or a problem — for example if a blood sample was not kept cold or was stored in the wrong tube.

Scientists need to take the lead on which QA elements they incorporate, says Melissa Eitzen, director of regulatory operations at the University of Texas Medical Branch in Galveston. “You want to give them tips that they can take or not take,” she says. “If they choose it, they’ll do it. If you tell them they have to do it, that’s a struggle.”

Rapport is paramount, says Michael Jamieson at the University of Southern California in Los Angeles, who helps other faculty members to move research towards clinical applications. Instead of talking about quality systems, he prefers to discuss concrete behaviours, such as labelling bottles with expiry dates and storage conditions. QA jargon puts scientists off, he says. “Using the term good research practice makes most researchers want to run the other way.”

It’s a lesson that many QA specialists have taken to heart. Some say ‘assessment’ or ‘quality improvement’ instead of ‘audit’. Even ‘research integrity’ can be an inflammatory phrase, says Davies. “You have to find a way to communicate that QA is not punitive or guilt-inspiring.”

NOT INTO TEMPTATION

Having data that are traceable — down to who did what experiment on which machine, and where the source data are stored — has knock-on benefits for research integrity, says Nett. “You can’t pick out the data that you want.” Researchers who must provide strong explanations about why they chose to leave any information out of their analysis will be less tempted to cherry-pick data. QA can also weed out digital meddling: popular spreadsheet programs such as Microsoft Excel can be vulnerable to errors or manipulation if not properly locked, but QA teams can set up instruments to store read-only files and prevent researchers from tampering with data accidentally or intentionally. “I can’t help but think that QA is going to make fraud harder,” says Davies.

And good quality systems can be contagious. Melanie Graham, who studies diabetes at the University of Minnesota, often collaborates with others to test potential treatments. More than once, she says, collaborators have sent her samples in a polystyrene tube with nothing but a single letter written on it. Graham sends it back and requests a label that specifies the sample’s identity and provenance, and a range of storage temperatures. ‘Keep frozen’ is too vague — she will not risk

SOLUTIONS

There are many things wrong with the fictitious lab shown on page 456. But, here are six that a quality-assurance manager would identify, and how they would solve them.



DISORGANIZED SAMPLE STORAGE

Clear labelling and proper organization are important for incubators and freezers. Everyone in the lab should be able to identify a sample, where it came from, who did what to it, how old it is and how it should be stored.



INADEQUATE DATA LOGGING

Data should be logged in a lab notebook, not scribbled onto memo paper or other detritus and carelessly transcribed. Notebooks should be bound or digital; loose paper can too easily be lost or removed.



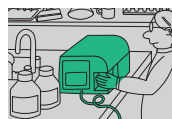
VARIABLE EXPERIMENTS

Protocols should be followed to the letter or deviations documented. If reagents need to be kept on ice while in use, each lab member must comply.



UNSECURED DATA ANALYSIS

Each lab member should have their own password for accessing and working with data, to make it clear who works on what, when. Some popular spreadsheet programs can be locked down so that manipulating data, even accidentally, is difficult.



MISSED MAINTENANCE

Instruments should be calibrated and maintained according to a regular, documented schedule.



OLD AND UNDATED REAGENTS

These can affect experimental results. Scientists should specify criteria for age and storage of all important reagents.

“I CAN’T HELP BUT THINK THAT QA IS GOING TO MAKE FRAUD HARDER.”

performing uninformative experiments because reagents stored in a standard freezer were supposed to be kept at -80°C .

When she first sent documentation requirements to collaborators, she expected them to push back. Instead, reactions were overwhelmingly positive. “It’s a relief for them,” says Graham. “They want us to handle their test article in a trusted way.”

The benefits go beyond providing solid data. In 2013, Davies worked with Torremorell and other Minnesota faculty members on a proposal to monitor and calibrate equipment used by several labs. The plan that they put in place helped them to secure US\$1.8 million to build shared lab space to deal with animal pathogens, says Torremorell. “If we want to be competitive to get funding, and if we want people to believe our data, we need to be serious about the data that we generate.”

Davies is still trying to spread the word. Her invitations to give talks and review grant applications have mushroomed. She and collaborators at other institutions have been developing online training materials and offering classes to technicians, postdocs, graduate students and principal investigators. After a presentation last year, a member of the audience told her that he had reviewed a grant from one of her clients; the QA plan had made the application stand out in a positive way. Davies was delighted. “I could finally come back to my folks and say, ‘It was noticed.’”

Davies knows it is still an uphill battle, but her ultimate goal is to make QA as much a part of research as peer review. It may not have the flash and dazzle of other efforts to ensure that research is robust and reproducible, but that is not the point. “A QA programme isn’t sexy,” says Michael Conzemius, a veterinary researcher at the University of Minnesota and another client of Quality Central. “It’s just kind of become the nuts and bolts of the scientific process for us.” ■

Monya Baker writes for *Nature* from San Francisco, California.

COMMENT

EVOLUTION Reflections on a best-seller, *The Selfish Gene*, at 40 **p.462**



HISTORY Mathematics was once mystical and practical **p.464**

COMMUNITY Concerns grow for Turkey's academics prosecuted by the state **p.466**

EMISSIONS A call to hold the atmosphere in trust and claim for damages **p.466**

ILLUSTRATION BY SÉBASTIEN THIBAUT



Don't let transparency damage science

Stephan Lewandowsky and Dorothy Bishop explain how the research community should protect its members from harassment, while encouraging the openness that has become essential to science.

Transparency has hit the headlines. In the wake of evidence that many research findings are not reproducible¹, the scientific community has launched initiatives to increase data sharing, transparency and open critique. As with any new development, there are unintended consequences. Many measures that can improve science² — shared data, post-publication

peer review and public engagement on social media — can be turned against scientists.

Endless information requests, complaints to researchers' universities, online harassment, distortion of scientific findings and even threats of violence: these were all recurring experiences shared by researchers from a broad range of disciplines at a Royal Society-sponsored meeting last year that we

organized to explore this topic. Orchestrated and well-funded harassment campaigns against researchers working in climate change and tobacco control are well documented^{3,4}. Some hard-line opponents to other research, such as that on nuclear fallout, vaccination, chronic fatigue syndrome or genetically modified organisms, although less resourced, have employed identical strategies. ►

► Such attacks place scientists in a difficult position. Good researchers do not turn away when confronted by alternative views. However, their openness can be exploited by opponents who are keen to stall inconvenient research. When people object to science because it challenges their beliefs or jeopardizes their interests, they are rarely committed to informed debate.

The progress of research demands transparency. But as scientists work to boost rigour, they risk making science more vulnerable to attacks. Awareness of tactics is paramount. Here, we describe ways to distinguish scrutiny from harassment.

USE AND ABUSE

We have identified ten red-flag areas that can help to differentiate healthy debate, problematic research practices and campaigns that masquerade as scientific inquiry (see 'Ten red flags'). None by itself is conclusive, but a preponderance of troubling signs can help to steer the responses of scientists and their institutions to criticism.

We also examine five legitimate tools of scholarly exchange, how they can be 'weaponized' (see 'Five doubled-edged tools' for a summary) and how to protect openness while curtailing its abuse.

Calls for open data: checking versus undermining. Many organized attacks call for more data, often with the aim of finding an analysis method that makes undesirable results go away⁵. The tobacco industry sponsored and drafted US legislation to enhance access to data on tobacco research, with the intention to delay or prevent evidence-based public-health measures⁶. Calls for more data can also be used to create the false impression that data are being withheld. In October last year, the chair of the Committee on Science, Space, and Technology of the US House of Representatives, a long-term critic of climate scientists, subpoenaed data from a federal agency that were already publicly available on the Internet (see go.nature.com/p4tmjd).

Protective action. We strongly support open data⁷, and scientists should not regard all requests for data as harassment.

When researchers cannot share data, they should explain why. Valid reasons may include confidentiality issues with clinical data, and cases in which participants' consent did not explicitly encompass data sharing. Researchers also need control over how data is to be used if it goes beyond what participants agreed to (for example, analysis of ethnic, race or gender differences in data collected for different purposes). The status of data availability should be enshrined in the publication record along with details about what information has been withheld and why. Some journals and publishers are already moving towards this practice (for example,

TEN RED FLAGS

Dr A publishes a study showing that food X increases the risk of disease Y. Critics accuse her of incompetence, scaremongering and ethical violations. Do these accusations constitute harassment or healthy debate?

	Raises red flags about researcher	Raises red flags about critics
Expertise	Does Dr A's contested work fall outside her training or her previous publications?	Are the critics operating outside their area of apparent expertise? Do the critics refuse to engage with the peer-reviewed literature?
Conflicts	Is Dr A funded by competitors of X? Is she marketing an antidote for Y?	Do the critics have a financial interest in the results?
Communication	Did Dr A promote this work without publishing it in a peer-reviewed journal?	Do the critics attack all researchers who show that X is harmful?
Errors	Does Dr A have a track record of major errors? Has she been defensive about minor errors?	Do the critics use small errors to dismiss all of Dr A's work?
Balance	Does Dr A have a record of misrepresenting evidence? Does she dismiss counter-arguments?	Do the critics have a record of cherry-picking evidence in public statements?
Scholarship	Are results out of line with existing, reputable scholarship, if it exists?	Can the critics specify what they would regard as convincing evidence?
Transparency	Has Dr A refused to make data available? Has she ignored reasonable disclosure standards?	Are the critics making showy demands for already-public data, or for data for which patients have not consented to publication?
Track record	Has Dr A routinely promoted flashy work without peer review?	Do the critics attack scientists across disciplines on different topics? Do they have a track record of harassment or vexatious complaints?
Insults or libel	Does Dr A uniformly dismiss critics as ignorant, biased or conflicted?	Are the critics levelling personal attacks? Are criticisms from anonymous sources or 'sock puppets'?
Freedom-of-information requests	Does Dr A claim that funding sources are irrelevant? Has she erected barricades to disclosure?	Do the critics use freedom-of-information requests for private correspondence unrelated to funding?

PLOS and some journals published by the Association for Psychological Science, including *Psychological Science*). Calls for a data set that ignore its open availability (including limitations agreed on during publication, where applicable) could suggest harassment.

We suspect that explicit discussion of what data are and are not available as part of the original publication process might have averted some of the ongoing controversy surrounding the PACE clinical trial, a UK study on chronic fatigue syndrome. The issue involves requests for data by transparency advocates, and the refusal by researchers and institutions to release data citing patient confidentiality, limited consent and requestors' intent.

Even when data availability is described in papers, tension may still arise if researchers do not trust the good faith of those requesting data, and if they suspect that requestors will cherry-pick data to discredit reasonable conclusions. Research is already moving towards study 'pre-registration' (researchers publishing their intended method and analysis plans before starting) as a way to avoid bias, and the same strictures should apply to critics during reanalysis. In general, critics and original researchers should obey symmetrical standards of openness and responsibility and be subject to symmetrical scrutiny

concerning conflicts of interest⁸. In cases in which researchers have no confidence in the good faith of the people requesting data, one potential solution would be arbitration by an independent adjudicator.

Social media: rapid correction versus mob rule. Blogs and social media enable rapid correction of science by scientists, as shown by the 'arsenic life' controversy in 2012, in which initial claims of a startling finding — that a bacterium could survive without phosphorus by substituting arsenic in its place in essential biomolecules — were rapidly rebutted by experts online (see *Nature* <http://doi.org/fx24wg>; 2012). Yet social media and online comments also offer an easy way to inject biased, incorrect or misleading information. And because engagement with critics is a core element of scientific practice, researchers may feel obliged to respond even to 'trolls' (online harassers).

Protective action. Scientists should ignore critics who are abusive or illogical and those that make the same points repeatedly despite rebuttals. Internet trolling has been associated with sadism and psychopathy⁹. Engagement with such bad-faith actors can imperil scientists' well-being in a way that university ethics committees would never condone in research on human subjects.

All who participate in post-publication review should identify themselves. The drawbacks of anonymity (its encouragement of bad behaviour) outweigh its advantages (for example, it allows junior people to criticize senior academics without fear of redress). What's more, the scientific community should not indulge in games of 'gotcha' (intentionally turning small errors against a person). Minor corrections and clarifications after publication should not be a reason to stigmatize fellow researchers. Scientific publications should be seen as 'living documents', with corrigenda an accepted — if unwelcome — part of scientific progress.

Freedom-of-information requests: right to know versus right to privacy. Freedom-of-information (FOI) requests have revealed conflicts of interest, including undisclosed funding of scientists by corporate interests such as pharmaceutical companies and utilities. But information requests have also been used as harassment, in attempts to embarrass researchers or just to waste their time. In 2010, the then-attorney-general of Virginia sought to obtain private e-mail correspondence from climate scientist Michael Mann, relating to work undertaken while he was at the University of Virginia in Charlottesville. This request, widely seen as a witch-hunt (see *Nature* **465**, 135–136; 2010), was ultimately struck down by the Virginia high court.

Protective action. Given that contemporary conversations are mainly conducted by e-mail, broad-ranging FOI laws risk being tantamount to permanent wiretaps in academics' offices. We fear that without the guarantee of privacy during e-mail conversations, self-censorship will have chilling effects on academic freedom and incisive discussion. A 2013 decision of the UK information commissioner towards preserving researchers' rights against disclosing "material which is still in the course of completion, to unfinished documents or to incomplete data" are encouraging, and cogent guidelines are beginning to emerge.

However, the right to privacy should not extend to funding arrangements¹⁰. Researchers should scrupulously disclose all sources of



Climate scientist Michael Mann was hounded for private e-mails.

funding; even small undisclosed amounts can create an impression of undue influence, as in a 2015 case involving a US researcher who was working on genetically modified crops and had received US\$25,000 from Monsanto to assist his outreach efforts (see *Nature* **524**, 145–146; 2015). FOI requests can be an appropriate tool in cases involving the conflation of public money and private interests.

Calls for retraction: correction versus censorship. Publication retractions have historically been reserved for cases of fraud or grave errors. Increasingly, however, calls for retraction are coming from people who do not like a paper's conclusions. In one famous case, a committee created by the National Football League called for a journal to retract an article by a medical researcher who argued that severe brain damage in a deceased American-football player had probably resulted from repeated concussions. (These conclusions were eventually endorsed by independent researchers.)

Protective action. Journals and professional societies should condemn specious calls for retraction. Journals and institutions can also publish threats of litigation, and use sunlight as a disinfectant.

Institutional self-scrutiny versus protection from harassment. Universities have complaint processes for good reasons. However, complaints are also used to undermine researchers doing legitimate but controversial science³.

Protective action. Scientists who are harassed often feel alone. Universities do not tolerate harassment based on race or gender, and neither should they tolerate harassment based on contentious science. They should provide training and support to help their researchers cope. Public declarations can be particularly useful: in 2014, in response to the harassment of one of its professors, the Rochester Institute of Technology in New York publicly acknowledged the scientific consensus on climate change and its support for academic freedom.

NEXT STEPS

Numerous professional bodies, educational institutions, government agencies and journals have convened meetings during the past few years to put science under the microscope. Issues such as reproducibility and conflicts of interest have legitimately attracted much scrutiny and have stimulated corrective action. As a result, the field is being invigorated by initiatives such as study pre-registration and open data.

Similar attention must be devoted to stressors and threats to science that arise in response to research that is considered inconvenient. The same institutions and bodies that have scrutinized science must also start a conversation about how to protect it. ■

Stephan Lewandowsky is professor in cognitive psychology at the University of Bristol, UK, who focuses on the public understanding of science. **Dorothy Bishop** is professor of developmental neuropsychology at the University of Oxford, UK; she chaired a symposium at the Wellcome Trust in London in April 2015 on improving scientific reliability. e-mails: stephan.lewandowsky@bristol.ac.uk; dorothy.bishop@psy.ox.ac.uk

FIVE DOUBLE-EDGED TOOLS

Legitimate tools of scholarly exchange can be weaponized.

Technique	Use	Abuse
Call for data	Permit the replication or inspection of analyses.	Impugn scientists' integrity (when data is already available); biased re-analyses.
Social-media posts	Highlight errors or questionable practices.	Stalk, libel, intimidate or harass.
Freedom-of-information requests	Reveal hidden conflicts of interest.	Launch a fishing expedition into private correspondence.
Call for retraction	Remove unethical or erroneous work from the literature.	Discredit inconvenient results.
Complaints to universities	Redress unethical conduct.	Damage reputation.

1. Nosek, B. A. *et al. Science* **348**, 1422–1425 (2015).
2. Alberts, B. *et al. Science* **348**, 1420–1422 (2015).
3. Landman, A. & Glantz, S. A. *Am. J. Public Health* **99**, 45–58 (2009).
4. Lewandowsky, S., Mann, M. E., Bauld, L., Hastings, G. & Loftus, E. F. 'The subterranean war on science' *APS Observer* (2013); available at <http://go.nature.com/iqtjsp>
5. Cataldo, J. K., Bero, L. A. & Malone, R. E. *J. Clin. Epidemiol.* **63**, 841–853 (2010).
6. Baba, A., Cook, D. M., McGarity, T. O. & Bero, L. A. *Am. J. Pub. Health* **95**, S20–S27 (2005).
7. Morey, R. D. *et al. R. Soc. Open Sci.* **2**, 15047 (2015).
8. Christakis, D. A. & Zimmerman, F. J. *J. Am. Med. Assoc.* **310**, 2499–2500 (2013).
9. Buckels, E. E., Trapnell, P. D. & Paulhus, D. L. *Personal. Individ. Diff.* **67**, 97–102 (2014).
10. Oreskes, N., Carlat, D., Mann, M. E., Thacker, P. D. & vom Saal, F. S. *Environ. Sci. Technol.* **49**, 7527–7528 (2015).



Richard Dawkins in 1976, around the time he published his first best-selling book.

IN RETROSPECT

The Selfish Gene

Matt Ridley reassesses Richard Dawkins's pivotal reframing of evolution, 40 years on.

Books about science tend to fall into two categories: those that explain it to lay people in the hope of cultivating a wide readership, and those that try to persuade fellow scientists to support a new theory, usually with equations. Books that achieve both — changing science and reaching the public — are rare. Charles Darwin's *On the Origin of Species* (1859) was one. *The Selfish Gene* by Richard Dawkins is another. From the moment of its publication 40 years ago, it has been a sparkling best-seller and a scientific game-changer.

The gene-centred view of evolution that Dawkins championed and crystallized is now central both to evolutionary theorizing and to lay commentaries on natural history such as wildlife documentaries. A bird or a bee risks its life and health to bring its offspring into the world not to help itself, and certainly not to help its species — the prevailing, lazy thinking of the 1960s, even among luminaries of evolution such as Julian Huxley and Konrad Lorenz — but (unconsciously) so that its genes go on. Genes that cause birds and bees to breed survive at the expense of other genes. No other explanation makes sense, although some insist that there are other ways to tell the story (see K. Laland *et al.* *Nature* **514**, 161–164; 2014).

What stood out was Dawkins's radical insistence that the digital information in a gene is effectively immortal and must be the primary unit of selection. No other unit shows such persistence — not chromosomes, not individuals, not groups and not species. These are ephemeral vehicles for genes, just as rowing boats are vehicles for the talents of rowers (his analogy).

As an example of how the book changed science as well as explained it, a throwaway remark by Dawkins led to an entirely new theory in genomics. In the third chapter, he raised the then-new conundrum of excess DNA. It was dawning on molecular biologists that humans possessed 30–50 times more DNA than they needed for protein-coding genes; some species, such as lungfish, had even more. About the usefulness of this “apparently surplus DNA”, Dawkins wrote that “from the point of view of the selfish genes themselves there is no paradox. The true ‘purpose’ of DNA is to survive, no more and no less. The simplest way to explain the surplus DNA is to suppose that it is a parasite.”

Four years later, two pairs of scientists published papers in *Nature* formally setting out this theory of “selfish DNA”, and acknowledged Dawkins as their inspiration (L. E. Orgel and F. H. C. Crick *Nature* **284**, 604–607 (1980); W. F. Doolittle and C. Sapienza *Nature* **284**, 601–603; 1980). Since then, Dawkins's

The Selfish Gene
RICHARD DAWKINS
Oxford University
Press: 1976.

TERRY SMITH/THE LIFE IMAGES COLLECTION/GETTY

speculation has been borne out by the discovery that much surplus DNA consists of reverse transcriptase — a viral enzyme whose job is to spread copies of itself — or simplified versions of transposons dependent on it. Thus, Dawkins's ideas helped to explain what was going on inside genomes, as well as between individuals, even though the book was written long before DNA sequencing became routine. The complexity of the structure of the gene itself has since grown enormously, with the discovery of introns, control sequences, RNA genes, alternative splicing and more. But the essential idea of a gene as a unit of heritable information remains, and Dawkins's synthesis stands to this day.

On *The Selfish Gene*'s 30th anniversary, many of Dawkins's admirers, including writer Philip Pullman and cognitive scientist Steven Pinker, contributed essays to the book *Richard Dawkins* (Oxford University Press, 2006) edited by his former students Alan Grafen and Mark Ridley (no relation of mine). In this Festschrift, the philosopher Daniel Dennett argued that the book was not just science, but “philosophy at its best”. In my contribution, I pointed out that the success of the book had spawned a gold rush for popular-science writers, as publishers began offering large advances in the hope of finding the next *Selfish Gene*. James Gleick's *Chaos* (Abacus, 1988), Stephen Hawking's *A Brief History of Time* (Bantam, 1988) and Pinker's *The Language Instinct* (William Morrow, 1994) were among the nuggets mined before the boom petered out.

Although his book brimmed with original thoughts, Dawkins was quick to acknowledge that he was building on the discoveries and insights of others, notably the evolutionary theorists William Hamilton, George Williams, John Maynard Smith and Robert Trivers. They were equally quick to appreciate that he had done something more than explain their ideas. Trivers wrote the foreword, and Maynard Smith narrated a television documentary about the book soon after it was published. Williams said in an interview that Dawkins's book had “advanced things a lot further than mine did” (see go.nature.com/21j1mt); Hamilton wrote that *The Selfish Gene* “succeeds in the seemingly impossible task of using simple, untechnical English to present some rather recondite and quasi-mathematical themes of recent evolutionary thought” in a way that would “surprise and refresh even many research biologists” (W. D. Hamilton *Science* **196**, 757–759; 1977).

As a first-year undergraduate in the zoology department at the University of Oxford, UK, where Dawkins was about to teach me

➔ **NATURE.COM**
For a review of
Richard Dawkins's
latest memoirs, see:
go.nature.com/cqukcq



Dawkins speaking at an atheist event in 2012.

computing and animal behaviour, I found the book exhilarating and bewildering. Until then, my teachers had helpfully divided the world into right ideas and wrong ones. But here was a writer turning some settled science upside down and inviting me to join him on a journey to discover a truth that seemed to him “stranger than fiction”. Was he right or wrong? I was being shown the arguments, not the answers.

The origin of *The Selfish Gene* is intriguing. Dawkins revealed in the first volume of his memoirs, *An Appetite for Wonder* (Bantam, 2013; see E. Scott *Nature* **501**, 163; 2013),

“Dawkins’s ideas helped to explain what was going on inside genomes long before DNA sequencing became routine.”

that the idea of selfish genes was born ten years before the book was published. In 1966, the Dutch biologist Niko Tinbergen asked Dawkins, then a research assistant with a new doctorate in animal behaviour, to give some lectures in his stead. Inspired by Hamilton, Dawkins wrote in his notes (reproduced in *An Appetite for Wonder*): “Genes are in a sense immortal. They pass through the generations, reshuffling themselves each time they pass from parent to offspring ... Natural selection will favour those genes which build themselves a body which is most likely to succeed in handing down safely to the next generation a large number of replicas of those genes ... our basic expectation on the basis of the orthodox, neo-Darwinian theory of evolution is that Genes will be ‘selfish.’”

Dawkins began writing the book in 1973,

and resumed it in 1975 while on sabbatical. At the suggestion of Desmond Morris, the zoologist and author of *The Naked Ape* (Jonathan Cape, 1967), Dawkins showed some draft chapters to Tom Maschler of Jonathan Cape, who strongly urged that the title be changed to ‘The Immortal Gene’. Today, Dawkins regrets not taking the advice. It might have short-circuited the endless arguments, so beloved of his critics and so redolent of the intentional stance (in which we tend to impute mental abilities to unconscious things, from thunderstorms to plants), about whether selfishness need be conscious. It might even have avoided the common misconception that Dawkins was advocating individual selfishness.

In the end, it was Michael Rodgers of Oxford University Press who enthusiastically published *The Selfish Gene*, after demanding “I must have that book!” when he saw early draft chapters. It was an immediate success, garnering more than 100 reviews, mostly positive. Dawkins went on to write books that were better in certain ways. *The Extended Phenotype* was more groundbreaking, *The Blind Watchmaker* more persuasive, *Climbing Mount Improbable* more logical, *River out of Eden* and *Unweaving the Rainbow* more lyrical, *The Ancestor’s Tale* more encyclopaedic, *The God Delusion* more controversial. But they were all variations on the themes he so eloquently and adventurously set out in *The Selfish Gene*. ■

Matt Ridley’s latest book is *The Evolution of Everything*. He is a columnist for *The Times*.

Twitter: @mattwridley



Pop-up diagrams in a 1570 translation of Euclid's *Elements*, for which John Dee wrote the preface.

HISTORY

Archive of wonders

Philip Ball browses remnants of the celebrated library of mathematician and occultist John Dee.

Before Galileo, it is difficult to tell the history of science as such. No one exemplifies that better than John Dee (1527–1609). The Tudor scholar was one of the most respected mathematicians of his day, advising explorers on navigation and fashioning ingenious mechanical devices for the theatre. He also cast horoscopes, stood accused of witchcraft, collected books of magic and professed to converse with angels. “Scholar, courtier, magician” is how he is described in an exhibition at the Royal College of Physicians (RCP) in London.

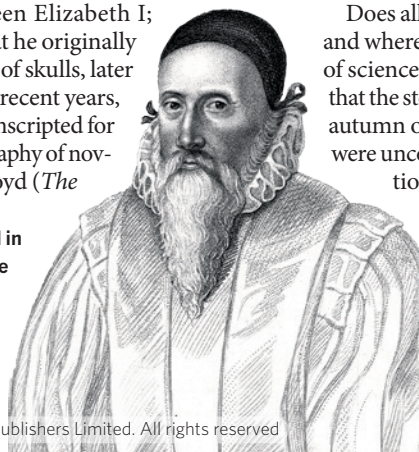
The show explores Dee's identity through the medium of his legendary library. At his home in Mortlake, west of London, Dee amassed perhaps the finest English library of his time, on topics from astronomy and alchemy to Greek poetry. Ultimately, the exhibition concedes that it is impossible either to sum up Dee's activities and interests under a single rubric or to understand them at all within the modern boundaries of science.

This is probably why Dee has enjoyed so many reincarnations in popular culture. Unlike the more clearly delineated Isaac Newton, we can refract him through the preoccupations of the age. Allegedly, he was the archetype for the magus Prospero in William Shakespeare's 1610 *The Tempest*, and he is said to have advised on the Globe Theatre's design. By 1659, when Dee's transcripts of angelic

conversations were posthumously published (after being exhumed; he had buried them in his garden), his work informed the debate about the reality of angels and demons. That was of interest a few years later to natural philosopher Robert Boyle, who believed that the existence of spirits vindicated religious ideas and undermined atheism. Boyle's colleague Robert Hooke claimed — unconvincingly — that Dee's angelic discourse contained encrypted intelligence for the court.

In 1806, these activities were adduced in a book on insanity and mass hysteria by English physician Thomas Arnold: supernatural phenomena now fell into the nascent realm of psychology, and Dee was seen as a deluded fanatic. A marvellous late-nineteenth-century painting by Henry Gillard Glindoni shows Dee conjuring with chemicals and fire before Queen Elizabeth I; X-rays show that he originally stood in a circle of skulls, later painted over. In recent years, Dee has been conscripted for the psychogeography of novelist Peter Ackroyd (*The*

John Dee pictured in an engraving made some 200 years after his death, by Robert Cooper.



Scholar, Courtier, Magician: The Lost Library of John Dee
Royal College of Physicians, London.
18 January – 29 July.

House of Doctor Dee; Hamish Hamilton, 1993) and postmodern fantasies of Albion in Damon Albarn's 2011 opera *Dr Dee*.

The RCP holds the largest remaining collection of Dee's books. Unsure of the English queen's favour, Dee left the country in 1583 under the patronage of disreputable Polish prince Olbracht Łaski, entrusting his library to his brother-in-law, Nicholas Fromond. It was a bad choice: Fromond allowed Dee's ‘friends’ and former pupils to carry off or buy the books. Dee returned, his fortunes having soured, to find most of his books gone. One energetic thief was Nicholas Saunder, possibly a former pupil: around 160 books that he filched ended up in the collection of Henry Pierrepont, the Marquis of Dorchester. They were among the 3,000 or so volumes bequeathed to the RCP around 1680, and several of them, supplemented by other books of Dee's from elsewhere, are now on display. They open a window on Dee's broad interests.

There are volumes on mining; Dee advised naval expeditions for voyages in search of far-flung gold to enrich Elizabeth's “British Empire” (a term that first appears in Dee's 1577 *Perfect Arte of Navigation*). He wrote the preface to a 1570 English translation of Euclid's *Elements*, complete with assemble-yourself pop-up figures of polyhedra and intersecting planes. Here, he defended mathematics against charges of witchcraft and explained its use in developing mechanical inventions and technologies. To Dee, mathematics was an art both mystical and practical: the key both to cosmic harmony and to something approaching quantitative science.

He read about navigation (and doodled a splendid galleon in the margin of Cicero's *Opera*), made notes on the weather, devised horoscopes in the margins of Girolamo Cardano's works on astrology. (He was arrested in 1555 on suspicion of illegally casting the horoscope of the reigning Catholic queen Mary, Elizabeth's predecessor.) Also on display is a fine collection of Dee's instruments, including an obsidian “magical mirror” brought from the New World and the original crystal ball in which his ‘scryer’ Edward Kelley claimed to see angels.

Does all this help us to make sense of Dee and where he fits (if at all) into the narrative of science? Rather, the exhibition shows us that the story has arbitrary boundaries in the autumn of the Renaissance, when scholars were unconstrained by disciplinary distinctions, and magic and marvels were still part of the rational cosmos. ■

Philip Ball is a freelance writer. His latest book is *Invisible*.
e-mail: p.ball@btinternet.com

Correspondence

Concern grows for Turkey's academics

We strongly urge the Turkish government to stop prosecuting academics, to abide by international human-rights values and to respect civil liberties — including freedom of speech (see *Nature* <http://doi.org/bbxj>; 2016).

In a petition to the government this month, more than 2,000 academics from Turkey and thousands of international scholars have called for an end to the curfews and violence against people in Kurdish provinces. This prompted Turkey's President Recep Tayyip Erdoğan to order the Higher Education Board to take action against those academics he described as committing “treason”. Istanbul's Chief Public Prosecutor launched a criminal investigation based on Article 301 of the Turkish Penal Code, which prosecutes those who insult the state.

We are deeply concerned about this escalating crisis. We hope that the international academic community will join us in condemning these attacks against our colleagues in Turkey.

Caghan Kizil* *German Centre for Neurodegenerative Diseases, Helmholtz Association, Dresden, Germany.*

caghan.kizil@crt-dresden.de

*Supported by 14 signatories (listed at go.nature.com/jy8ol6).

Synthesize evidence to steer decisions

Using evidence mapping to display and categorize environmental studies cannot replace ‘evidence synthesis’ in guiding decision-making (M. C. McKinnon *et al. Nature* **528**, 185–187; 2015). There are no shortcuts to evidence-based practice.

The results of investigations need to be synthesized to allow conclusions to be drawn from contradicting data (L. V. Dicks *et al. Trends Ecol. Evol.* **29**,

607–613; 2014). Studies can be assigned a ‘level of evidence’ indicator of design and quality, which is derived from evidence hierarchies (see, for example, A.-C. Mupepele *et al. Ecol. Appl.* <http://dx.doi.org/10.1890/15-0595.1>; 2016). This indicator reflects the confidence with which the reported outcome can be causally attributed to the investigated driver.

Practitioners' questions are rarely answered directly by an existing set of studies. Evidence-based medicine tackles this problem by developing clinical guidelines on the basis of collated scientific results and clinical experience, and by using systematic reviews of research results and evidence assessments that are supported by hierarchies.

Anne-Christine Mupepele,
Carsten F. Dormann *University of Freiburg, Germany.*
anne-christine.mupepele@biom.uni-freiburg.de

Hold atmosphere in trust for all

We, the undersigned, call on the V20 — the 20 countries that are most vulnerable to the effects of climate change — to take the lead in creating an ‘atmospheric trust’ that establishes community property rights over the atmospheric commons (www.claimthesky.org). The V20 could use this trust as a legal instrument to address the climate crisis and to help to implement last month's Paris agreement to keep warming well below 2 °C.

Under public-trust doctrine, certain natural resources such as soil and water must be held in trust to serve the public good. It is every government's responsibility as a trustee to protect these assets as natural capital and to maintain them for the public's use, not give them away or sell them to private parties. The global atmosphere is one such asset.

An atmospheric trust would act as an independent agency and trustee. It could collect

claims for damages to the atmosphere and invest funds in mitigation, adaptation and compensation, and in resources for the most affected populations. Because only 90 enterprises (mainly extractive industries) are responsible for two-thirds of global carbon emissions (R. Heede *Clim. Change* **122**, 229–241; 2014), damage claims could target a relatively small number of entities.

All governments would eventually be co-trustees in the atmospheric commons, with a fiduciary responsibility to protect it from catastrophic releases of greenhouse gases.

Robert Costanza* *The Australian National University, Acton, Australia.*

robert.costanza@anu.edu.au

*On behalf of 31 correspondents (see go.nature.com/52f8mt for full list).

What stops women getting more grants?

Women make up 33% of the applicants who are eligible for programmes funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC), but they lead only 21% of grant applications. The percentage receiving large grants of more than £2 million (US\$2.8 million) remains stubbornly low: in 2014, women had a success rate of 17% compared with 44% for men. To investigate this, we informally surveyed focus groups from seven BBSRC-funded universities (see go.nature.com/wqrfz3).

All groups cited society's expectations of professional women and (unconscious) biases against them. They also specified the way in which science as a profession organizes itself and how esteem is rewarded; and dominant behaviours by full-time researchers (primarily men) that seem to attract support at the expense of more-junior, part-time or flexitime researchers (primarily women). There were perceived inconsistencies

in the grant-award process, including in the quality and tone of reviewer comments and committee feedback, and concerns about gender imbalance in the reviewer pool.

The BBSRC is working with its research communities to address these issues. We welcome suggestions that could help us to achieve a more diverse and inclusive research community (see also M. Urry *Nature* **528**, 471–473; 2015).

David McAllister, Jan Juillerat, Jackie Hunter *BBSRC, Swindon, UK.*

david.mcallister@bbsrc.ac.uk

Solar energy needs focus

The high cost of solar photovoltaic installations prevents them from providing more than about 1% of the world's electricity requirement. A solution would be to incorporate an optical concentrator in the solar photovoltaic module that would save on expensive materials without compromising electrical output.

Optical concentrators focus solar energy on a small area attached to a photovoltaic cell (P. Gleckman *et al. Nature* **339**, 198–200; 1989). However, this technology has been held back by its complex manufacturing and assembly processes, its modest electrical-conversion efficiency and a lack of government funding and policy.

Researchers, industries and governments must work together to resolve the technical issues associated with this promising technology and come up with a practical, industry-ready design to revive the solar energy market.

Abu Bakar Munir *University of Malaya, Malaysia.*

Firdaus Muhammad-Sukki *Robert Gordon University, Aberdeen, UK; and Multimedia University, Selangor, Malaysia.*

Nurul Aini Bani *Universiti Teknologi Malaysia, Malaysia.*
f.b.muhammad-sukki@rgu.ac.uk

GENOME EDITING

The domestication of Cas9

The enzyme Cas9 is used in genome editing to cut selected DNA sequences, but it also creates breaks at off-target sites. Protein engineering has now been used to make Cas9 enzymes that have minimal off-target effects. [SEE ARTICLE P.490](#)

FYODOR URNOV

For the past 30,000 years, humans have been genetically engineering the wolf through selective breeding, preserving some forms of genes and eliminating others to produce the dog. Now, two studies (one on page 490 of this issue¹ and one in *Science*²) have used genetic engineering to tame a different type of wild creature — a nuclease enzyme called Cas9. In doing so, they have markedly reduced the enzyme's undesirable natural tendencies, but have preserved its ability to cut DNA in an RNA-guided manner. This feat of molecular domestication is great news for practitioners of genome editing, in which DNA sequences in cells or organisms are changed to scientists' specifications efficiently and accurately; such precise editing requires highly targetable nucleases³.

In its natural 'wild' state, Cas9 is part of the bacterial immune system. When a bacterium is infected by a parasite such as a virus, the organism's cellular machinery cuts up and retains pieces of the invader's DNA, storing the sequences in a region of the bacterium's own genome called a CRISPR locus⁴. Cas9 then polices the bacterium for repeat invaders by carrying with it an RNA copy of a sequence stored in the CRISPR locus (for simplicity, this RNA is referred to here as a guide RNA, or gRNA). The enzyme compares intracellular DNA to the sequence in the gRNA, and if there is a match, Cas9 cuts the invading DNA. Attackers evade detection by changing their DNA sequence, so Cas9 evolved to cut incoming DNA even if its sequence is a less-than-perfect match to the gRNA.

Studies of this and other bacterial defence mechanisms have had a major impact on genome editing. In this process, a nuclease cuts DNA inside the cell and, as this DNA break is being repaired, the desired edit (disruption, correction or insertion of a gene) takes place^{3–5}. The first genome-editing experiments made use of another class of nuclease, zinc-finger nucleases (ZFNs)⁵, but the discovery that Cas9 is led by a gRNA dramatically expanded the

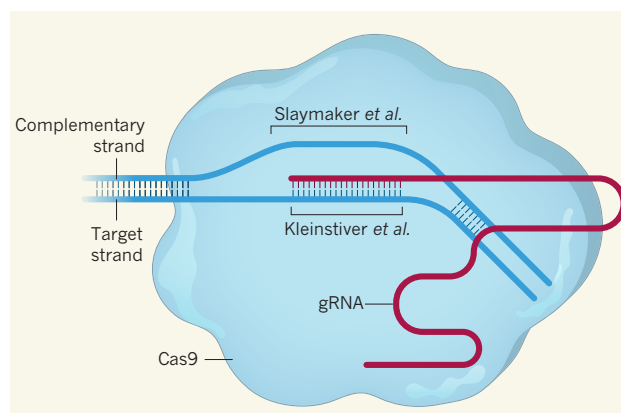


Figure 1 | Taming a wild enzyme. The Cas9 enzyme cuts specific DNA sequences, which it identifies using a guide RNA (gRNA) that pairs with the chosen sequence in an unwound DNA double helix. Kleinstiver *et al.*¹ engineered Cas9 such that interactions between the enzyme and the backbone of the gRNA-paired DNA were weakened. Slaymaker *et al.*² engineered the contacts that the enzyme makes with the complementary DNA single strand, which is not recognized by the gRNA. These modifications forced the engineered enzymes to rely to a greater extent on the gRNA for sequence recognition, thus improving their binding specificity. (Adapted from ref. 1.)

scale and scope of genome-editing applications for research purposes⁶, because the gRNA enables easy and relatively efficient enzyme programming.

Cas9 evolved to defend a bacterium that has a genome 1,800 times smaller than the human genome, and cutting DNA sequences that are imperfectly matched to the gRNA is an adaptation to its natural battlefield. As a consequence, when Cas9 was brought in from the wild and placed in human cells, it introduced genetic changes to unintended stretches of DNA in addition to editing the gene of interest⁷. Imagine a short-sighted witness to a crime attempting to identify the perpetrator in a police line-up, relying not only on the facial features that make the criminal unique but also on those shared with other people, such as gender or height. This could lead to a case of mistaken identity, because a weak match to the witness's fuzzy mental image could be reinforced by a match on shared features. Similarly, wild-type Cas9 finds its target not only using the sequence-specific gRNA, but also by grasping onto the DNA backbone, which is the same in any gene.

In the present studies, Kleinstiver *et al.*¹ and Slaymaker *et al.*² set out to tame Cas9 using a thoughtful and well-executed approach that relied on an atom-by-atom understanding of how the enzyme binds to and cuts DNA⁸. (A similar study of how zinc fingers bind DNA⁹ ultimately provided the basis for the first genome-editing experiments.) The groups reasoned that, by engineering Cas9 such that its interactions with the DNA backbone were weakened, they could force the enzyme to rely to a greater extent on the gRNA–DNA pairing to recognize and cut its target (Fig. 1).

Kleinstiver and colleagues tested the editing specificity of their resulting enzyme, which they dubbed high-fidelity Cas9 (Cas9-HF), in a cancer-cell line. They programmed Cas9-HF with gRNAs for seven different stretches of human DNA: in six cases it edited only the intended target, and in the remaining stretch

the enzyme was weakly distracted by only one other position in the DNA. By comparison, wild-type Cas9 cut at multiple unintended sequences when tested with gRNAs for these seven gene targets. Crucially, for 75% of targets tested, Cas9-HF was just as potent a genome editor as its wild ancestor. Slaymaker and colleagues followed the same overall principle to produce an enzyme that they called enhanced specificity Cas9 (eCas9), although the details of their engineering and analyses differed.

These 'domesticated' Cas9 enzymes are sure to be used by laboratories the world over. The immediate impact will be to shorten the time it takes to complete a genome-editing experiment, because the need to check for undesired edits will be reduced. Cas9 has been used to systematically scan many human genes at a time for those that underlie a trait of interest¹⁰, and such experiments will now be more efficient. In agriculture — especially in species that have extended life cycles, such as crops or cattle — the use of enhanced Cas9 might obviate the need to perform time-consuming crosses to obtain a pristinely edited organism.

Genome editing was first applied in the clinic

in 2009, when an *ex vivo* ZFN-based approach was used to edit certain immune cells from people with HIV¹¹. This approach has since been used to treat more than 80 patients, and has a good safety record. Last December, the first clinical trial of *in vivo* gene editing, a ZFN-based approach for treating haemophilia¹², passed review by the US Food and Drug Administration. ZFNs have already been engineered to a level of specificity that is comparable to that of Cas9-HF (see go.nature.com/mkl6v1), and they have passed regulatory hurdles for use in clinical trials in both *in vivo* and *ex vivo* applications. The current studies inspire

confidence that the scope of clinical genome editing will continue to expand. Advances in this field offer the promise of engineering genetic cures for many diseases — a prospect that is both encouraging and within our reach. ■

Fyodor Urnov is at Sangamo BioSciences, Richmond, California 94804, USA.
e-mail: furnov@sangamo.com

1. Kleinstiver, B. P. *et al.* *Nature* **529**, 490–495 (2016).
2. Slaymaker, I. M. *et al.* *Science* **351**, 84–88 (2016).
3. Carroll, D. *Annu. Rev. Biochem.* **83**, 409–439 (2014).
4. Doudna, J. A. & Charpentier, E. *Science* **346**, 1258096 (2014).

5. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. *Nature Rev. Genet.* **11**, 636–646 (2010).
6. Bolukbasi, M. F., Gupta, A. & Wolfe, S. A. *Nature Methods* **13**, 41–50 (2015).
7. Tsai, S. Q. *et al.* *Nature Biotechnol.* **33**, 187–197 (2015).
8. Jiang, F. & Doudna, J. A. *Curr. Opin. Struct. Biol.* **30**, 100–111 (2015).
9. Pavletich, N. P. & Pabo, C. O. *Science* **252**, 809–817 (1991).
10. Shalem, O., Sanjana, N. E. & Zhang, F. *Nature Rev. Genet.* **16**, 299–311 (2015).
11. Tebas, P. *et al.* *N. Engl. J. Med.* **370**, 901–910 (2014).
12. Sharma, R. *et al.* *Blood* **126**, 1777–1784 (2015).

The author declares competing financial interests: see go.nature.com/vvxzp6 for details.

NEUROSCIENCE

Fluorescent boost for voltage sensors

The development of a voltage sensor in which a microbial rhodopsin protein is fused with a fluorescent protein enables the neuronal activity of single cells in live animals to be measured with unprecedented speed and accuracy.

VIVIANA GRADINARU
& NICHOLAS C. FLYTZANIS

Over the past few decades, scientific discoveries have greatly deepened our understanding of biology, but our ability to understand the workings of our own minds has proved a frustrating exception. It is therefore no surprise that a major aim of neuroscience is to develop tools, in particular light-based technologies, with which to deconstruct brain function and dysfunction by controlling and recording brain activity¹. Writing in *Science*, Gong *et al.*² report one of the best tools so far for the optical monitoring of neuronal activity — a strongly fluorescent, fast-acting, voltage-sensing protein that can be used to image subcellular activity changes in model organisms.

The brain uses complex electrical signalling that is orders of magnitude more efficient than the fastest computer to react to external and internal stimuli, processing these inputs and outputting behaviours. The action potentials that make up this signalling are caused by the opening and closing of protein channels in the cell membrane through which ions can flow into or out of the cell. Changes in relative ion concentrations alter the voltage across the cell membrane, causing rapid propagation of electrical currents down the length of the neuron and eventual signalling to other neurons downstream. To decode this neuronal language, we need tools that can track electrical activity from neuron to neuron across the brain.

Genetically encoded calcium indicators

(GECIs)³, which fluoresce in response to the calcium-ion influx that is triggered by neuronal activity, are used as standard for tracking electrical activity in neurons *in vivo*. These sensor proteins can be used to monitor both population-wide and single-cell activity in chosen cell types. However, influx of

calcium ions is an indirect and slow measure of action potentials and cannot capture all smaller, subthreshold events — changes in membrane voltage that are not large enough to trigger an action potential, but that can nonetheless affect brain physiology.

An alternative approach involves fluorescent proteins called genetically encoded voltage indicators (GEVIs), which directly report electrical activity. This approach has undergone major developments^{1,4} in the past few years, but, until now, could not detect fast neuronal activity in live animals. The ability to read and discriminate across the entire bandwidth of action-potential frequencies is key, because specific activities underlie distinct behaviours. Furthermore, some cells fire at high rates, and slow sensors such as GECIs might not discriminate such cells from their slower neighbours.

Gong *et al.* developed a GEVI that can

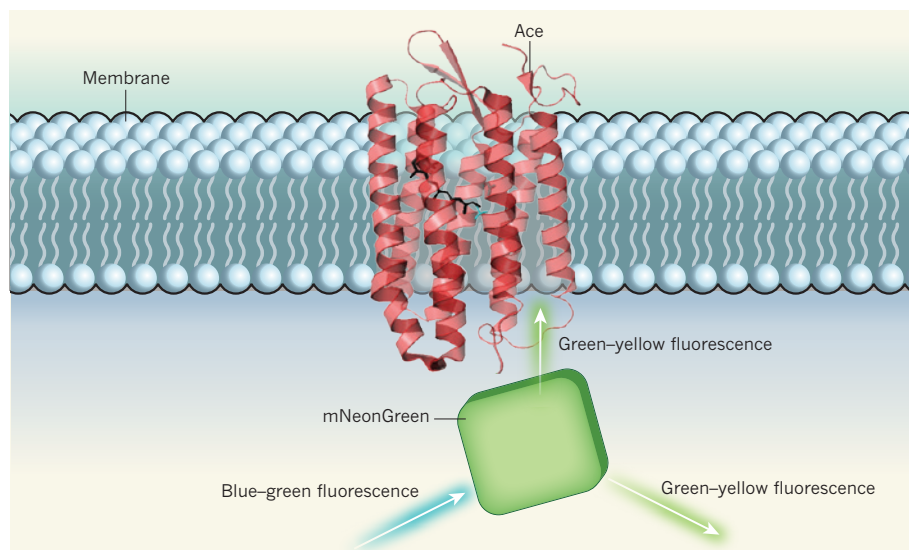


Figure 1 | Sensor with a light touch. Gong *et al.*² have developed a genetically encoded voltage sensor that rapidly and accurately detects neuronal activity in live animals. The sensor is comprised of a microbial rhodopsin protein, Ace, fused to a fluorescent protein, mNeonGreen. mNeonGreen is excited by blue-green light and emits green-yellow fluorescence. Ace spans the cell membrane and absorbs green-yellow light in a voltage-dependent manner, with more light absorbed during action potentials, when the voltage increases across the membrane. Thus, the overall level of green-yellow light emitted by the fusion protein provides a readout for electrical activity, with higher light emission indicating lower membrane voltage or inactivity. (Adapted from the Gradinaru Group, California Institute of Technology.)

monitor fast neuronal activity in single cells in live flies and mice. The authors used a microbial rhodopsin protein called Ace as the basis for their voltage sensor. Microbial rhodopsins are light-sensitive ion channels, and were initially adopted in neuroscience for their ability to generate electrical currents and so to modulate neuronal activity¹. More recently, these proteins have been used to monitor electrical currents because they fluoresce in a voltage-dependent manner⁵. They are fast and sensitive sensors, but their use in live organisms⁶ has been hampered by the fact that they fluoresce only weakly.

The researchers bypassed this obstacle by fusing Ace with the fluorescent protein mNeonGreen. In this configuration, blue-green light excites mNeonGreen, which emits green-yellow fluorescence. A portion of this fluorescence is absorbed in a voltage-dependent manner by Ace, causing mNeonGreen-emitted fluorescence to decrease as the membrane voltage rises and neuronal activity increases, and to increase as the membrane voltage falls (Fig. 1). *In vitro*, the Ace-mNeon fusion protein acts six times faster and can resolve closely spaced, repeating action potentials much more accurately than similar protein fusions⁷.

To assess the capabilities of their tool *in vivo*, Gong and colleagues compared it with GECIs in live mice and flies. Measurements taken using Ace-mNeon during a visual task corroborated previous measurements taken with GECIs. In mice, Ace-mNeon flawlessly reported single action potentials in neurons at the surface of the brain's cortex region, 20 times faster than is possible using GECIs. This is an impressive achievement, because intact mammalian tissue is opaque and can be naturally fluorescent — both of which are factors that can mask the signal from fluorescent proteins.

In flies, Ace-mNeon recorded more than 18,000 action potentials with perfect accuracy, and detected odour-evoked subthreshold and fast voltage changes that a GECI failed to pick up. Furthermore, the authors used the protein to track voltage propagation from one side of a cell to the other with submillisecond precision. Such precision tracking was previously unachievable in live flies.

Although the sensor's performance is impressive, major challenges remain before it can replace GECIs *in vivo*. First, the authors used conventional fluorescence microscopy for *in vivo* imaging. The effectiveness of this type of imaging for sensor detection relies on sparse expression of Ace-mNeon, limiting the number of cells that can be imaged concurrently. Second, for maximum impact, a fast sensor requires fast imaging, but imaging speed and field of view are inversely correlated in current imaging techniques, so rapid imaging limits the ability to simultaneously investigate many cells. The combination of fluorescence microscopy and limited field of view meant that Gong *et al.* could study only a handful of cells

at a time. A third challenge is that, although mNeonGreen is three times more stable to light than other rhodopsin-paired fluorescent proteins, extended continuous imaging sessions still 'bleach' the protein, decreasing its fluorescence. This limitation could be bypassed by using multiple short exposures, or by spacing measurements widely enough for protein turnover to replace the photobleached sensors.

The benefits of using GEVIs such as Ace-mNeon to image activity in live animals are undeniable. Nonetheless, better hardware is required to realize the full potential of these voltage reporters. Until that is available, calcium sensors will remain the gold standard for studying densely labelled cell populations simultaneously over extended imaging sessions, especially in deep brain areas. The development of technologies such as microendoscopy⁸ and fibre photometry⁹ has enabled calcium imaging of subcortical brain regions, and fine-tuning these techniques for use with GEVIs is an exciting possibility for the future. Overall, Gong and colleagues' study

highlights the power of microbial rhodopsins, especially when paired with strongly fluorescent proteins, and the need for continued development of these tools hand-in-hand with microscopy techniques. ■

Viviana Gradinaru and Nicholas C. Flytzanis are in the Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA.

e-mail: viviana@caltech.edu

- Emiliani, V., Cohen, A. E., Deisseroth, K. & Hausser, M. *J. Neurosci.* **35**, 13917–13926 (2015).
- Gong, Y. *et al.* *Science* **350**, 1361–1366 (2015).
- Chen, T.-W. *et al.* *Nature* **499**, 295–300 (2013).
- Knöpfel, T., Gallero-Salas, Y. & Song, C. *Curr. Opin. Chem. Biol.* **27**, 75–83 (2015).
- Kralj, J. M., Douglass, A. D., Hochbaum, D. R., MacLaurin, D. & Cohen, A. E. *Nature Meth.* **9**, 90–95 (2012).
- Flytzanis, N. C. *et al.* *Nature Commun.* **5**, 4894 (2014).
- Gong, Y., Wagner, M. J., Li, J. Z. & Schnitzer, M. J. *Nature Commun.* **5**, 3674 (2014).
- Lecoq, J. *et al.* *Nature Neurosci.* **17**, 1825–1829 (2014).
- Lerner, T. N. *et al.* *Cell* **162**, 635–647 (2015).

EVOLUTION

A lizard that generates heat

Birds and mammals generate heat to regulate body temperature, but most non-avian reptiles cannot. The discovery of endothermy during the reproductive period of a tegu lizard sheds light on the evolution of this characteristic.

COLLEEN G. FARMER

The avian and mammalian lineages diverged 320 million years ago, and since that time both lineages have converged on a radically different approach to life from that of their common ancestor. Birds and mammals are endotherms, meaning they use internal heat to regulate their body temperature; their ancestor, and many extant animals such as amphibians and non-avian reptiles, are ectotherms that rely on external heat sources (Fig. 1). Understanding the convergent evolution of endothermy in birds and mammals is a central question in evolutionary physiology, because thermal biology is linked to fundamental traits such as body size, food requirements and aspects of reproduction. Writing in *Science Advances*, Tattersall *et al.*¹ report the remarkable discovery that a lizard species uses endothermy during its reproductive period. Their finding supports the idea^{2,3} that the ability to exert control over temperature during reproduction was the common selective agent that drove the evolution of

endothermy in birds and mammals.

The transition from aquatic to terrestrial habitats presented animals with new challenges to reproduction; chief among these was the fact that eggs laid on land are at risk of desiccation and are subject to greater fluctuations in temperature than eggs laid in water. One lineage of animals — the amniotes — evolved eggs containing a series of fluid-filled membranes, which reduced the risk of desiccation (Fig. 1). Many amniotes further evolved an ability to exert control over temperature during reproduction. For example, viviparity (giving birth to live young rather than laying eggs) allows females to control developmental temperature by gaining heat through basking, and has evolved independently more than 100 times in lizards and snakes⁴.

Tattersall *et al.* studied black and white tegu lizards (*Salvator merianae*), which inhabit tropical, subtropical and temperate climates throughout the plains east of the Andes Mountains. During autumn and winter, the lizards hibernate in burrows, after which their reproductive phase begins. Males undergo a surge in

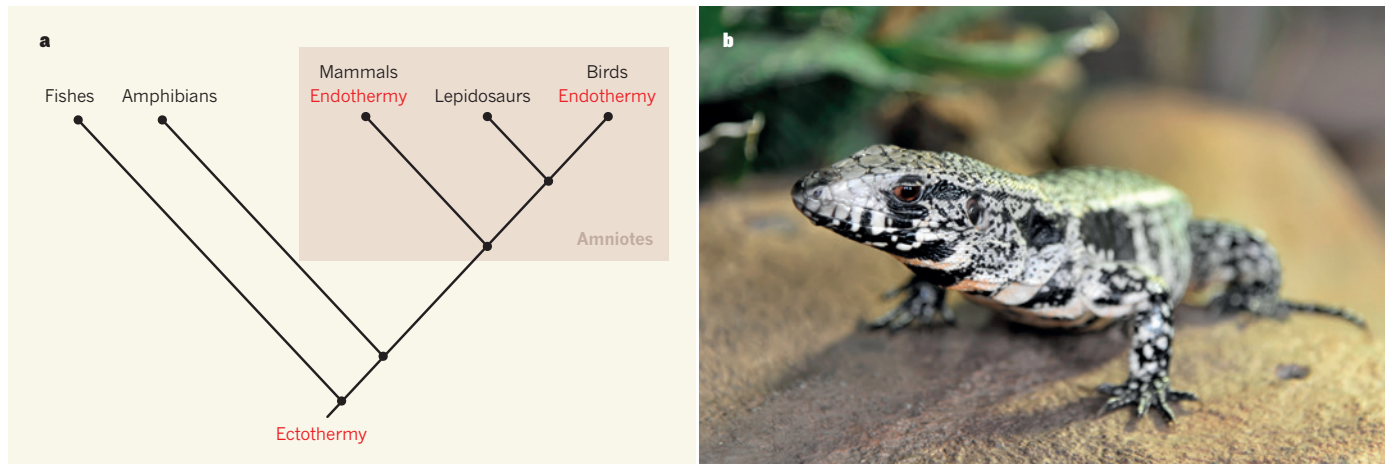


Figure 1 | Thermal strategies during reproduction. **a**, Most fishes and aquatic amphibians are ectotherms, which rely on external sources of heat. These animals lay many small eggs in water, exerting no parental control of temperature during reproduction. That strategy changed with the evolution of amniotes — animals whose eggs have fluid-filled membranes that allow development on land. Mammals and birds have both evolved endothermy, meaning they generate body heat, which they can use to incubate eggs

(monotremes and birds) or retained embryos (marsupials and placentals). By contrast, lepidosaurs, which include lizards and snakes, are ectotherms that mostly exert reproductive temperature control through egg incubation after basking or by retaining embryos through to live birth. **b**, Tattersall *et al.*¹ report that black and white tegu lizards (*Salvator merianae*), which are ectotherms during most life stages, use endothermy during their reproductive period.

testosterone and gonadal growth, emerge from their burrows and establish territories, but initially forgo foraging. After the reproductive period, they reduce activity, feed heavily and gain weight^{5,6}. When females end hibernation, they mate and deposit yolk in their eggs, which entails a heavy energy investment — clutch mass is typically about 40% of body mass^{7,8}. Clutches are laid in nests made of various materials, including moist grass, small sticks and other litter⁶, which probably improves the insulating properties of the nest.

After laying, the females remain with the eggs for up to around 75 days⁶ with little or no foraging activity. Female attendance greatly influences nest temperature; one study found attended nests to be 5 °C warmer than a control nest where females were barred from brooding⁹. Because these lizards are capital breeders — that is, their reproduction is decoupled temporally from food acquisition and assimilation — changes in body temperature during their reproductive period cannot be explained by an increase in metabolism associated with feeding.

Tattersall *et al.* investigated the relationship between reproduction and thermoregulation in sexually mature tegu lizards reared in a captive colony. The body temperatures of both male and female lizards were equal to their burrow temperatures during most of the hibernation period, except for between days 160 and 180, during which the researchers recorded an increase in body temperature above burrow temperature. This is the period in which the lizards rouse from hibernation and begin the reproductive period. The lizards supplemented their endogenous heat production by basking to gain heat during the day, retreating to their burrows at night. Remarkably, body temperature remained elevated throughout the night, whereas during the non-reproductive

season, body temperature equilibrated with the temperature of the burrow.

With these observations, Tattersall *et al.* have established that, during the reproductive period and when insulated by a burrow, these relatively small (around 2-kilogram) lizards can generate heat that raises their body temperature by up to 10 °C above ambient, and that this thermogenesis is not related to feeding or activity. Furthermore, the observations refute conventional wisdom that small animals lacking body insulation, such as hair and feathers, cannot significantly increase their body temperature.

The authors also placed reproductive-phase, fasting lizards in a temperature-controlled chamber for 8 days, and found that they maintained body temperatures that were greater than ambient. Disturbing the lizards caused their body temperatures to decline, possibly owing to increased heat dissipation as a result of elevated peripheral blood flow. This observation may explain why endothermy has been missed by other researchers, who have measured body temperature in disturbed animals rather than quiet, undisturbed animals.

Tattersall and colleagues' work not only provides the first evidence of endothermy in a lizard, but also complements previous findings of endothermy during reproduction in pythons¹⁰. Like the tegu lizards, diamond pythons (*Morelia spilota*) construct insulated nests and achieve body temperatures of up to 13 °C above ambient when brooding¹¹. We now know that reproductive endothermy is not an oddity of one clade of snakes. Indeed, there is increasing evidence that many species of bird and mammal improve their capacities for thermogenesis and endothermy during reproduction (reviewed in refs 2, 3).

The selective drivers for the evolution of endothermy are debated. However, convergent

evolution is one of the strongest lines of evidence for the adaptive significance of a trait. Thus, this discovery in lizards corroborates the idea that the initial selective benefits of the evolution of endothermy in birds and mammals were reproductive^{2,3}. Studies of pythons have shown that the thermal regime during incubation affects the incubation period as well as the characteristics of the hatchlings (such as initial growth rates, escape behaviour and willingness to feed¹²), providing several potential bases on which reproductive endothermy may provide an evolutionary advantage.

Intriguing questions remain. How do these lizards generate body heat, and do so only at certain times? Precisely how does thermogenesis facilitate the lizards' reproduction — might it expand the geographical range over which this species can reproduce, or alter the time window for reproduction? Are tegus and pythons alone, or are there other reproductively endothermic non-avian reptiles? Reproductive endothermy may yet be discovered in other species if they are studied using methods that do not disturb them during the reproductive period, when insulating nests reduce rates of heat dissipation and metabolism is increased by the synthetic demands of reproduction. ■

Colleen G. Farmer is in the Department of Biology, University of Utah, Salt Lake City, Utah 84112-0840, USA.
e-mail: cg.fmr@gmail.com

1. Tattersall, G. J. *et al.* *Sci. Adv.* **2**, e1500951 (2016).
2. Farmer, C. G. *Am. Nat.* **155**, 326–334 (2000).
3. Farmer, C. G. *Am. Nat.* **162**, 826–840 (2003).
4. Shine, R. *Annu. Rev. Ecol. Evol. Syst.* **36**, 23–46 (2005).
5. Winck, G. R. & Cechin, S. Z. *J. Nat. Hist.* **42**, 239–247 (2008).
6. Fitzgerald, L. A., Chani, J. M. & Donadio, O. E. in *Neotropical Wildlife Use and Conservation* (eds Robinson, J. & Redford, K. H.) 303–316

(Univ. Chicago Press, 1991).

7. Manes, M. E., Noriega, T., Casal, C. & Apichela, S. *Cuad. Herpetol.* **21**, 21–29 (2007).
8. Lopes, H. R. & Abe, A. S. in *Manejo y Conservación de Fauna Silvestre en América Latina* (eds Fang, T. G.,

Montenegro, O. L. & Bodmer, R. E.) 259–272 (Inst. Ecol., La Paz, 1999).

9. Manes, M. E., Ibanez, M. A. & Manilla, A. *Rev. Argentina Prod. Anim.* **23**, 119–126 (2003).
10. Hutchison, V. H., Dowling, H. G. & Vinegar, A. *Science*

11, 694–695 (1966).

11. Slip, D. J. & Shine, R. *J. Zool.* **216**, 367–378 (1988).

12. Shine, R., Madsen, T. R. L., Elphick, M. J. & Harlow, P. S. *Ecology* **78**, 1713–1721 (1997).

INFECTION BIOLOGY

Small RNA with a large impact

A simultaneous comparison of the RNA molecules expressed by *Salmonella* bacteria and human cells during infection reveals how a bacterial small RNA alters the transcript profiles of both the bacteria and the host cells. [SEE ARTICLE P.496](#)

MATTHIAS P. MACHNER & GISELA STORZ

What happens when bacteria encounter or enter host cells? How does each of the species respond, the bacteria to survive in their new environment and the host cells to either tolerate non-harmful bacteria or defend against pathogenic ones? To answer these questions, it is imperative to understand how gene transcription in both cells changes during the encounter. Over the years, approaches applied to this problem have ranged from *in vivo* gene-expression technology¹ to sequencing the full complement of bacterial or host-cell transcripts^{2,3} (the transcriptome). However, such analyses have largely focused on messenger RNAs and have profiled either the bacteria or the host, not both at once. In this issue,

Westermann *et al.*⁴ (page 496) go beyond the individual organisms by using dual RNA-seq, an approach that simultaneously profiles bacterial and host transcriptomes throughout the course of an infection.

The RNA-seq method takes advantage of the ever-increasing depth of sequencing (the number of reads for a particular sample) now possible. Westermann *et al.* first assessed whether the dual RNA-seq approach accurately reflected known gene regulation in human HeLa cells and in the bacterium *Salmonella enterica* serovar Typhimurium (hereafter *Salmonella*), a common cause of food poisoning, during infection. The authors' data confirmed that, as previously reported⁵, transcription of invasion-related genes in the genomic region known as *Salmonella* pathogenicity island 1 (SPI-1) was reduced after

bacterial internalization, whereas transcription of SPI-2 genes, which promote intracellular survival, increased.

Having validated the sensitivity of the approach, Westermann *et al.* focused on mRNAs and regulatory RNAs whose expression changed during the course of the 24-hour infection. In bacteria, small regulatory RNAs (sRNAs) that base-pair with target mRNAs to modulate the mRNA's stability or translation are integral to a wide range of stress responses, including the response to host cells⁶. Thus, the authors were intrigued by an 80-nucleotide sRNA, which they denoted PinT, whose expression was highly induced during infection, and which was activated by the bacterial PhoP/Q system, known to be crucial for *Salmonella* survival in the intracellular environment.

A striking finding of the dual RNA-seq analysis was that tens of bacterial and hundreds of host-cell transcripts were affected merely by the presence or absence of PinT (Fig. 1). On the bacterial side, overproduction of PinT led to reduced levels of the mRNAs encoding SopE and SopE2, two SPI-1 effector proteins that mediate host-cell invasion by *Salmonella*. These mRNAs were elevated in strains lacking the *pinT* gene. By mutating the *pinT*, *sopE* and *sopE2* sequences, the authors revealed that the inhibitory effect of PinT occurred through direct base-pairing with the mRNAs. Dual RNA-seq also revealed a role for PinT in repressing SPI-2 genes later in infection. However, control of these genes was indirect and occurred through PinT base-pairing with the mRNA that encodes the cyclic AMP receptor protein (CRP), an activator of transcription of SPI-2 genes. These data indicate that PinT, on bacterial internalization, controls the temporal expression of both SPI-1 effectors and SPI-2 virulence genes, thus facilitating the bacterium's transition from an invasive state to a state of intracellular replication.

Westermann *et al.* then compared the transcriptomes of the host HeLa cells challenged with either wild-type *Salmonella* or a strain lacking PinT. They discovered numerous changes in cells infected with the PinT-lacking mutant, including altered levels of many long non-coding RNAs (lncRNAs), hyperactivation of mitochondrial genes, increased abundance of mRNAs for proteins involved in innate immune pathways (such as the interleukin-8 mRNA) and accelerated activation of SOCS3, a protein that regulates the inflammatory JAK–STAT signalling pathway. The last finding is of particular interest, because properly balanced JAK–STAT signalling is essential for

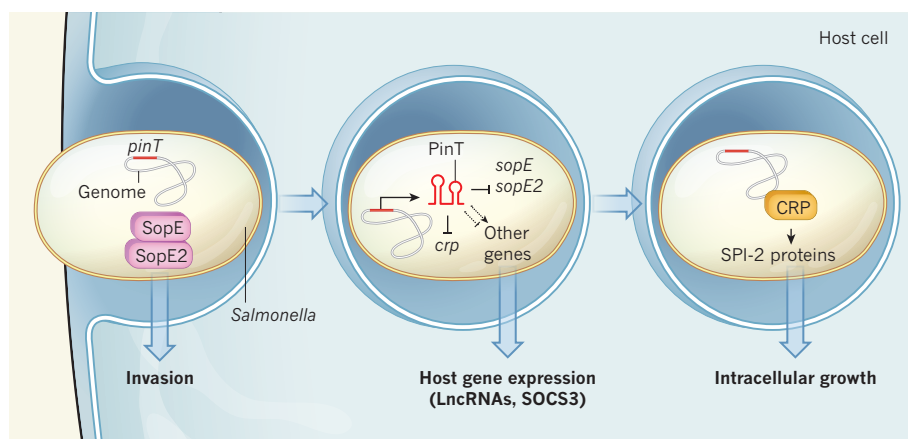


Figure 1 | PinT orchestrates gene expression in *Salmonella* and its host cells. When *Salmonella* bacteria invade cells, expression of the genes *sopE* and *sopE2* facilitates bacterial invasion. After the bacteria are internalized, the levels of the transcripts from these genes fall. By simultaneously monitoring the RNA molecules present in both *Salmonella* and host cells over the course of an infection, Westermann *et al.*⁴ found that a small regulatory RNA expressed in *Salmonella*, which they name PinT, induces this repression by base-pairing with the *sopE* and *sopE2* messenger RNAs. PinT also base-pairs with the mRNA that encodes CRP, a protein that activates transcription of genes encoding SPI-2 proteins. This repression is reduced later in infection, allowing the SPI-2 proteins to regulate the bacterium's intracellular growth. The authors also observed differences in host-cell transcripts when the cells were infected with *Salmonella* mutants lacking PinT, including altered levels of long non-coding RNAs (lncRNAs) and the mRNA for SOCS3. This suggests that PinT targets other bacterial genes that influence host-cell gene expression.

optimal *Salmonella* infection. Too little inflammation reduces the ability of *Salmonella* to compete with the intestinal microbiota, whereas too much inflammation might result in the bacteria being killed by the host^{7–9}.

The dual RNA-seq approach of simultaneously interrogating the bacteria and host transcriptomes gives unprecedented insight into the dynamic RNA-expression landscape during the bacterium–host interaction. This method is particularly useful for analysing genes such as *pinT* that, when deleted, cause little or no detectable change in standard bacterial virulence assays, but nevertheless have a strong impact on mRNA levels. The authors' study of PinT, however, also demonstrates some of the ongoing challenges in studying bacterium–host interactions. For example, although the effect of PinT on host gene expression such as JAK–STAT signalling is interesting, it remains unclear which PinT-controlled bacterial gene(s) are responsible. In addition, it has become evident that there is substantial variation in gene expression

even between individual bacteria in single-species cultures; these differences are not captured by approaches in which cell samples are sequenced in bulk.

Finally, although dual RNA-seq eliminates the need to separate bacteria from the host cell for investigation, a comprehensive transcriptome analysis over the course of an infection still demands the analysis of a series of time points in short succession. Westermann *et al.* conducted a time series for their analysis of HeLa cells infected with wild-type and PinT-lacking *Salmonella*, but this approach is impractical for the analysis of large collections of bacterial mutants or isolates. Continued improvement in deep-sequencing technologies, including single-cell sequencing, will undoubtedly circumvent some of these limitations. In the meantime, the application of Westermann and colleagues' dual RNA-seq approach to a wide range of bacterium–host interactions will open a treasure trove of insight into transcriptional actions and reactions as bacteria enter and proliferate in host cells. ■

Matthias P. Machner and Gisela Storz are in the Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland 20892, USA. e-mails: machnerm@mail.nih.gov; storzg@mail.nih.gov

1. Slauch, J. M., Mahan, M. J. & Mekalanos, J. J. *Meth. Enzymol.* **235**, 481–492 (1994).
2. Mandlik, A. *et al. Cell Host Microbe* **10**, 165–174 (2011).
3. Xiu, L. *et al. Genet. Mol. Res.* **14**, 16948–16965 (2015).
4. Westermann, A. J. *et al. Nature* **529**, 496–501 (2016).
5. Hautefort, I. *et al. Cell. Microbiol.* **10**, 958–984 (2008).
6. Wagner, E. G. H. & Romby, P. *Adv. Genet.* **90**, 133–208 (2015).
7. Uchiya, K. & Nikai, T. *Infect Immun.* **73**, 5587–5594 (2005).
8. Hannemann, S., Gao, B. & Galán, J. E. *PLoS Pathog.* **9**, e1003668 (2013).
9. Stecher, B. *et al. PLoS Biol.* **5**, 2177–2189 (2007).

This article was published online on 20 January 2016.

STELLAR ASTROPHYSICS

The mystery of globular clusters

The discovery of multiple stellar populations — formed at different times — in several young star clusters adds to the debate on the nature and origin of such populations in globular clusters from the early Universe. [SEE LETTER P.502](#)

ANTONELLA NOTA & CORINNE CHARBONNEL

Since the discovery of the first globular cluster in 1665, these large, ancient agglomerates of stars — which can host up to a million suns — have fascinated both astronomers and the public. They are visible through small telescopes, and their exquisite spherical symmetry singles them out in the sky and makes them easy to classify. However, their formation and evolution history is unclear. On page 502 of this issue, Li *et al.*¹ report observations of young star clusters that may help to crack the mystery of the oldest star clusters.

Globular clusters remain gravitationally bound as they orbit their host galaxy, on a timescale comparable to the lifetime of the low-mass stars they host. They are 10 billion to 13 billion years old — their age defines the boundaries of the age of the Universe. These well-studied systems were long thought to be simple and to host a single population of stars that all formed at the same time.

But in 2004, everything we knew about globular clusters changed radically. Using

accurate Hubble Space Telescope photometry^{2,3}, astronomers detected not one, but multiple stellar populations in ω Centauri, one of the most massive globular clusters in the Milky Way. Subsequent studies (see ref. 4, for example) of other globular clusters confirmed that this was not an isolated finding but the discovery of a general feature that revolutionized our understanding of such objects. These stellar populations were shown to exhibit unique chemical properties that are not found in any other stellar environment⁵. This means that these clusters are not simple at all, and have experienced more than one star-forming event during their lifetime.

The exciting news^{2,3} inspired different star-formation models to account for the photometric and spectroscopic properties of the different populations hosted by single clusters. For example, colliding winds from late-stage, medium-mass stars or ejecta from fast-rotating massive stars were invoked as a trigger and/or an origin of the processed material that fuelled the second generation of star formation. Astronomers also proposed an explanation based on a single generation of stars. They

suggested that very-low-mass stars with protoplanetary disks — gas rotating around newly formed stars — could acquire the observed properties by sweeping up material shed from interacting binary or rapidly rotating massive stars in a continual growth process.

However, most of these theories suffer from major issues⁶. Some of them imply a more massive original cluster than observed, with a substantial fraction of the original stars lost to the galactic halo. But this enhanced mass is in contrast to theoretical expectations for the dynamical evolution of these systems⁷ and to observations of dwarf galaxies and of young massive clusters in our Local Group of galaxies⁸. The consensus in the community is that we urgently need alternative, innovative ideas to overcome the impasse.

Enter Li and colleagues¹. The authors challenge the status quo by presenting observations of three massive clusters that are 1 billion to 2 billion years old in the Magellanic Clouds, members of our Local Group. The authors show clear evidence of a late burst of star formation that occurred a few hundred million (up to one billion, within the errors) years after the clusters' initial formation epoch. Multiple stellar population sequences are visible even in the authors' raw diagrams. The colours of the younger stellar sequences are consistent with an enhanced abundance of helium. This would be expected as a result of the chemical anomalies in globular clusters that are due to hydrogen burning at high temperature (helium is the main yield of hydrogen burning).

To explain the data, Li *et al.* propose that such clusters orbiting within the gaseous disks of their host galaxies could accrete sufficient gas reservoirs to form the next generation of stars. They suggest that this mechanism may

account for the ubiquitous multiple stellar populations in globular clusters, assuming that the observed massive young star clusters in the Magellanic Clouds are the modern counterparts of the old globular clusters. It is a plausible working hypothesis, but not a consensus view: some astronomers think that the formation mechanism of young massive star clusters in our neighbouring galaxies might be different from the globular-cluster formation that occurred in the early Universe. The link between young and old clusters has yet to be fully established. The determination of whether the chemical properties of the youngest stars in the Magellanic clusters are similar to those of globular-cluster stars will eventually address the issue.

Li and colleagues may not yet have all the answers needed to put the proposed general model on a firm quantitative footing. For example, the mass in the younger observed

population sequence is much smaller than that typically observed in globular clusters. There is no explanation given for the origin of the additional helium that is claimed to be present in the youngest stars, nor for the fact that the accreted material must have the same metal content as the original proto-cluster, even though the host galaxies might have evolved chemically between the distinct star-formation events. (By 'metal', astronomers mean any element after helium in the periodic table.)

Nevertheless, the findings present an innovative approach that deserves further attention. It will certainly advance the ongoing debate, as well as trigger original thoughts, future observations and corresponding interpretations. And it could lead to a final, robust explanation in the not too distant future — an example of how scientific debate works at its best. ■

Antonella Nota is at the Space Telescope Science Institute and European Space Agency, Baltimore, Maryland 21218, USA.

Corinne Charbonnel is at the Geneva Observatory, University of Geneva, CH-1290 Versoix, Switzerland, and at the Institut de Recherche en Astrophysique et Planétologie, Toulouse, France.

e-mails: nota@stsci.edu;

corinne.charbonnel@unige.ch

1. Li, C. *et al.* *Nature* **529**, 502–504 (2016).
2. Bedin, L. R. *et al.* *Astrophys. J.* **605**, L125–L128 (2004).
3. Piotto, G. *et al.* *Astrophys. J.* **621**, 777–784 (2005).
4. Milone, A. P. *et al.* *Astrophys. J.* **808**, 51 (2015).
5. Gratton, R. G., Carretta, E. & Bragaglia, A. *Astron. Astrophys. Rev.* **20**, 50 (2012).
6. Bastian, N. Preprint at <http://arxiv.org/abs/1510.01330> (2015).
7. Krause, M. G. H., Charbonnel, C., Bastian, N. & Diehl, R. Preprint at <http://arxiv.org/abs/1512.04256> (2015).
8. Larsen, S. S., Brodie, J. P., Grundahl, F. & Strader, J. *Astrophys. J.* **797**, 15 (2014).

ions (K^+) and magnesium ions. Strikingly, although oxygen and glucose deprivation did cause an inward flow of ions in oligodendrocytes that was mediated by glutamate release, this current flow seemed to be triggered by an increase in extracellular K^+ concentration and closure of K^+ -channel proteins, rather than by glutamate receptors.

Similar to an earlier report⁵, Hamilton and colleagues found that ischaemia led to an increase in intracellular Ca^{2+} levels in oligodendrocytes and their axon-ensheathing processes. However, inhibition of NMDA and AMPA/kainate receptors did not prevent Ca^{2+} influx. These data suggest that glutamate receptors are not the primary channels responsible for the ischaemia-evoked elevations of Ca^{2+} concentrations that cause myelin damage.

Next, the authors demonstrated that the ischaemia-evoked elevation of extracellular K^+ levels leads to increased hydrogen-ion levels (acidification) in oligodendrocytes that in turn triggers Ca^{2+} influx (Fig. 1). By increasing local intracellular H^+ levels in oligodendrocytes and measuring intracellular Ca^{2+} changes, Hamilton *et al.* investigated which other channels might contribute to ischaemia-evoked Ca^{2+} influx. After taking into account the known physiological properties of different ion-channel proteins and testing the effects of channel stimulators and inhibitors, the authors concluded that the channel responsible is TRPA1 — a widely expressed member of the family of transient receptor potential (TRP) channels (Fig. 1). Activation of TRPA1 channels allows Ca^{2+} , Mg^{2+} and Na^+ to enter the cell.

In line with Hamilton and colleagues' conclusion, ischaemia-triggered Ca^{2+} entry was considerably reduced in white-matter cerebellum slices from mice in which TRPA1 had been deleted. However, the authors did not test the *in vivo* effects of ischaemia on white matter in these mice. Instead, using isolated optic nerves

NEUROSCIENCE

A mechanism for myelin injury

The cells that insulate neuronal processes with a myelin membrane sheath are damaged during stroke. Data now show that an influx of calcium ions mediated by the TRPA1 protein contributes to myelin injury. [SEE LETTER P.523](#)

AIMAN S. SAAB & KLAUS-ARMIN NAVE

Normal brain function requires the rapid transmission of information between brain regions along neuronal projections called axons. The ability of axons to conduct information depends on the well-being of a supporting class of glial cells called oligodendrocytes¹, which speed up conduction by enveloping the axonal projections in a multilayered membrane sheath called myelin. Damage to oligodendrocytes and the myelin sheaths that they produce has been associated with axonal dysfunction in numerous disorders, including cerebral palsy, spinal-cord injury, multiple sclerosis and stroke. In this issue, Hamilton *et al.*² (page 523) report that the mechanisms that underlie this damage are more complex than commonly thought and involve the activation of a channel protein called TRPA1.

During a stroke, the local loss of blood flow in the brain, known as ischaemia, causes damage to neurons and glial cells, including oligodendrocytes. Even transient ischaemia causes permanent defects in axonal conductivity that can be only partially restored when the supply of oxygen and glucose to the tissue is

re-established³. Ischaemia causes the release of the neurotransmitter molecule glutamate, which excites oligodendrocytes. Some evidence^{3–5} indicates that blocking glutamate-receptor proteins reduces myelin damage and axon dysfunction during ischaemia. AMPA/kainate-type and NMDA-type glutamate receptors are ion-channel proteins that, when activated, allow positively charged ions such as sodium (Na^+) and calcium (Ca^{2+}) to flow into the cell. High levels of Ca^{2+} are toxic to cells, and so the death of oligodendrocytes and injury to myelin sheaths during ischaemia are widely thought to reflect the overactivation of these glutamate receptors^{5–7} — the same mechanism by which ischaemia damages neurons.

Hamilton *et al.* revisited this issue in brain slices from the rat cerebellum. They found evidence that, in the cerebellar white matter (a tissue that contains a high density of myelinated axons) deprivation of oxygen and glucose — a model of ischaemia-evoked oligodendrocyte and myelin damage — might not result in glutamate-receptor overactivation alone. First, the authors characterized ischaemia-evoked membrane currents in oligodendrocytes and monitored the corresponding intracellular changes in Ca^{2+} , Na^+ , potassium

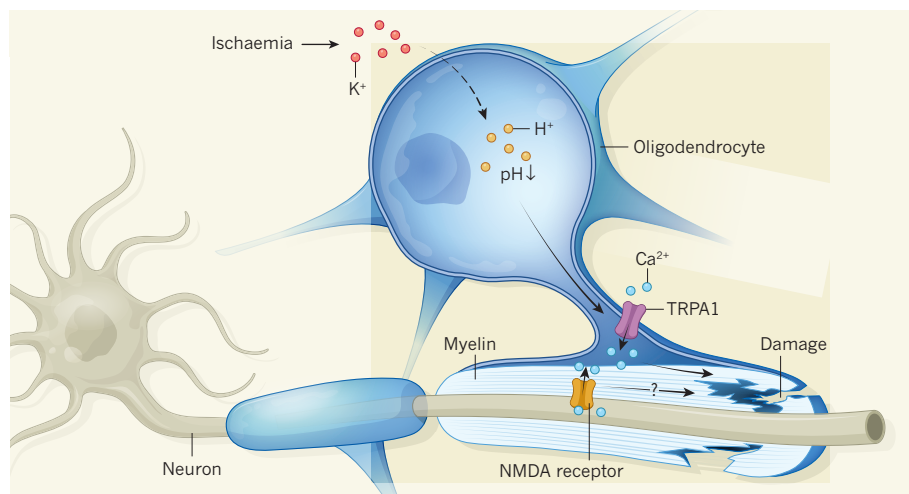


Figure 1 | Myelin injury in a model of ischaemia. Oligodendrocyte cells wrap around neurons and produce an insulating stack of membranes called myelin that speeds up neuronal signal transmission. When brain regions are deprived of glucose and oxygen (a condition called ischaemia), oligodendrocytes and myelin become damaged. Hamilton *et al.*² report that this damage is caused, in part, by a pathway that involves an increase in extracellular levels of potassium ions (K^+). Through an unknown mechanism (dashed arrow), increased K^+ levels trigger an intracellular rise in hydrogen ions (H^+). This reduces the pH in the cell, activating TRPA1-channel proteins and leading to an influx of calcium ions (Ca^{2+}). High levels of Ca^{2+} are toxic to oligodendrocytes and damage myelin. Glutamate-receptor proteins such as NMDA-type receptors can also mediate Ca^{2+} influx, but whether they have a role in ischaemia-evoked myelin damage in this setting is unclear.

from rats, the authors showed that blocking TRPA1 channels during oxygen and glucose deprivation reduced myelin damage, but had no effect on axonal injury. Thus, unlike the case with oligodendrocytes, damage to axons is mediated not by the activation of TRPA1 channels, but by other mechanisms of Ca^{2+} influx.

Hamilton *et al.* also detected spontaneous Ca^{2+} -level changes in the myelin of some normal oligodendrocytes, which were not deprived of glucose and oxygen. This spontaneous activity is probably indicative of axonal activity and the concurrent release of glutamate. So, given that myelin contains AMPA/kainate and NMDA receptors^{5–7}, and that ischaemia-evoked Ca^{2+} -level changes in optic nerves have previously been shown to be caused by activation of glutamate receptors⁵, why did the authors not find evidence for glutamate-evoked Ca^{2+} changes?

The levels of the messenger RNAs that encode NMDA receptors are low in mature oligodendrocytes, but the levels of TRPA1 mRNA are almost undetectable⁸, and the researchers still detected TRPA1-mediated Ca^{2+} influx in cerebellar oligodendrocytes. Evidence indicates that NMDA receptors move to myelin processes that face the axonal surface as oligodendrocytes mature^{5,6}. This might help to explain why glutamate-evoked Ca^{2+} responses are difficult to detect in oligodendrocytes. Alternatively, perhaps oligodendrocytes in the optic nerve, which myelinate axons of only glutamate-releasing neurons, are different from the cerebellar oligodendrocytes that also myelinate neurons producing a different neurotransmitter, GABA.

In contrast to the current paper, a study published last month⁹ provided evidence that,

in optic nerves, NMDA and AMPA/kainate receptors do indeed act to mediate Ca^{2+} influx in mature myelin. Such influx in myelin is mediated by axonal activity and the vesicular release of glutamate. However, beyond the level of Ca^{2+} signals, the physiological functions of oligodendroglial NMDA receptors still need to be resolved.

Hamilton and colleagues' study demonstrates that *in vitro* damage to myelin owing to oxygen

and glucose deprivation is more complex than anticipated, with TRPA1 channels joining the scene. Whether the results apply *in vivo* or to human TRPA1 channels, which have a different pharmacological response from those of mice¹⁰, remains to be seen. Trials of NMDA-receptor blockers in people who have had a stroke have largely failed, in part because the drugs were administered too late, given the low doses at which their adverse side effects can be tolerated. Identifying other pharmacological targets raises the hope that safer drugs may be found. ■

Aiman S. Saab and Klaus-Armin Nave are at the Max Planck Institute of Experimental Medicine, Department of Neurogenetics, 37075 Göttingen, Germany. A.S.S. is also at the Institute of Pharmacology and Toxicology, University of Zurich, and at the Neuroscience Center Zurich, Zurich, Switzerland. e-mails: asaab@pharma.uzh.ch; nave@em.mpg.de

1. Nave, K.-A. *Nature* **468**, 244–252 (2010).
2. Hamilton, N. B., Kolodziejczyk, K., Kougiumtzidou, E. & Attwell, D. *Nature* **529**, 523–527 (2016).
3. Tekkok, S. B., Ye, Z. & Ransom, B. R. *J. Cereb. Blood Flow Metab.* **27**, 1540–1552 (2007).
4. Follett, P. L., Rosenberg, P. A., Volpe, J. J. & Jensen, F. E. *J. Neurosci.* **20**, 9235–9241 (2000).
5. Micu, I. *et al.* *Nature* **439**, 988–992 (2006).
6. Salter, M. G. & Fern, R. *Nature* **438**, 1167–1171 (2005).
7. Kárádóttir, R., Cavalier, P., Bergersen, L. H. & Attwell, D. *Nature* **438**, 1162–1166 (2005).
8. Zhang, Y. *et al.* *J. Neurosci.* **34**, 11929–11947 (2014).
9. Micu, I. *et al.* *Exp. Neurol.* **276**, 41–50 (2015).
10. Nagatomo, K. & Kubo, Y. *Proc. Natl Acad. Sci. USA* **105**, 17373–17378 (2008).

This article was published online on 13 January 2016.

BIOANALYTICAL DEVICES

Technological leap for sweat sensing

Sweat analysis is an ideal method for continuously tracking a person's physiological state, but developing devices for this is difficult. A wearable sweat monitor that measures several biomarkers is a breakthrough. SEE LETTER P.509

JASON HEIKENFELD

Athletics trainers, physicians and even local pharmacists can take a sample of blood, saliva or urine and measure a whole panel of analytes (dissolved compounds) to reveal your physiological status at the time of sample collection. But none of the measurement techniques involved is conveniently portable or can continuously collect data for many hours or days — with the exception of glucose monitoring, which typically

requires blood samples to be drawn by needle at regular intervals. On page 509 of this issue, Gao *et al.*¹ report a truly non-invasive, continuous biomonitoring device: a wearable, Bluetooth-enabled band containing a panel of sensors for sodium, potassium, lactate, glucose and skin temperature. And rather than using the body fluids mentioned above, the device measures analytes in human sweat.

Making a wearable band that electrochemically senses sweat analytes is extremely difficult. The sensors must be prepared from

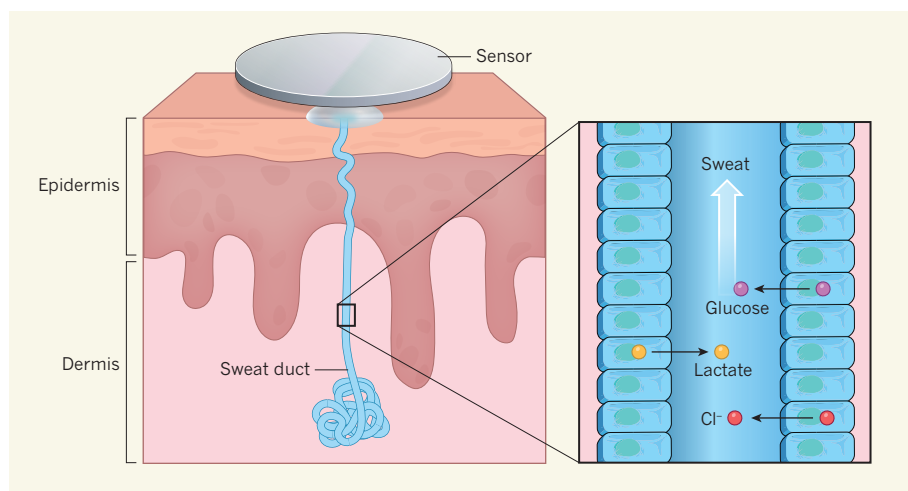


Figure 1 | Analysing sweat. Many biomarkers for a subject's physiological state, such as glucose, lactate or chloride ions (Cl^-), enter sweat from cells that form the walls of sweat ducts inside the skin. Gao *et al.*¹ report devices that can be worn as wrist or head bands, and which continuously analyse several molecules and ions in sweat using sensors placed on the skin's surface. (Adapted from ref. 2.)

scratch from basic chemicals — they can't just be purchased like the accelerometer chips used in smart watches and activity trackers. Another challenge is creating electronics that work with the ultra-high electrical impedance of the sensors. Basically, you need to figure out how to take a potentiostat — a device used to control electroanalytical experiments that typically weighs more than 2 kilograms — and make it so small and thin that you can wrap it around your wrist.

For decades, sweat analysis was relegated mainly to medical labs; this hampered its broader use for two reasons. The first is obvious: who can afford to tote around a cadre of trained medical staff and the associated equipment? The second reason is that conventional clinical methods for sweat collection and sensing could lead to inaccurate measurements. This is because existing clinical infrastructure is ill-equipped to work with the tiny volumes obtainable from sweat.

Gao *et al.* address these problems by putting tiny electronic sweat sensors right up against the skin (Fig. 1) — an approach that others have also reported^{2–6}. Sweat and its analytes are thus quickly measured as they emerge onto the skin's surface. These sensors are highly electrochemically selective³, and, despite their miniature sizes (on the order of square millimetres or smaller), they can distinguish a single type of ion or molecule from thousands of others in sweat.

This ability is a real leap forward for wearable devices, and couldn't have been made just by improving the rudimentary electrical or optical sensors found in commercially available activity trackers. For example, commercial trackers at best use a simple measure of electrical conductance on skin as a non-quantitative measure of sweat rate, whereas measuring sodium and potassium concentrations with electrochemical sensors quantifies sweat rate²

and could also quantify the total amount of electrolytes lost during exercise.

Importantly, Gao and colleagues' devices use many sensors. Previous devices have been limited to a single sensor, which could generate misleading information — if a stand-alone sensor shows a signal change, it could be because sweating has stopped, because the sensor has fallen away from the skin, or even because the sensor is failing. Having multiple sensors can clarify what is happening. For example, potassium levels in sweat are fairly invariant with sweat rate and with normal physiological changes in the body². So if there is a change in sodium, lactate or glucose signals while the potassium signal holds steady, then the other sensor changes can be trusted to be caused by a real physiological event.

The Bluetooth capability of their devices enabled Gao *et al.* to monitor continuously recorded data for at least an hour, and the types of sensors and electronics used should, in theory, enable such monitoring for 24 hours or more. Previously reported devices lacked Bluetooth. Having this capability is certainly commercially relevant, and start-up companies have developed functional, but unpublished, Bluetooth sweat-sensing technology in the form of watches⁷ or patches⁸.

The potential applications of wearable sweat-sensing devices extend well beyond those related to exercise. For example, the hormone cortisol is a marker of stress, and its concentrations in sweat are similar to those found in blood², making it a possible target for future monitoring. Even small-molecule drugs and their metabolites come out in sweat, so this body fluid might one day be used to monitor the amount of active drug in a patient's blood — helping to avoid rises and falls in drug levels between doses.

Today's commercially available wearables largely rely on decades-old technology. Their

market success is due to a convergence of improved affordability and ergonomics and a rapidly growing consumer awareness of health. The next watershed in wearables will probably be driven by scientific breakthroughs. Sweat biomonitoring arguably has the greatest potential among the emergent non-invasive technologies. But such potential will go unfulfilled unless scientists dig into the unresolved 'hard science' of this approach.

For example, cutting-edge, commercial, point-of-care blood sampling and sensing technologies proudly claim that as little as 20 microlitres of blood are needed for some tests. But square-millimetre-sized sensors on the skin will receive, at most, several nanolitres of sweat per minute². Just placing such sensors against the skin does not fully resolve this problem, because the gap between a sensor and the rough surface of the skin is so large that it takes periods of tens of minutes for fresh samples of sweat to displace previously accumulated sweat². Although not exactly real-time monitoring for your workout, this a good start, and is certainly better than repeated blood draws.

Consider also possible applications involving situations in which you are unlikely to be sweating, such as monitoring your medication levels while at the office. Methods exist for locally stimulating sweat by iontophoresis — that is, using a tiny electrical current to drive a chemical sweat stimulant into the skin. But these methods were commercialized for collecting single sweat samples, not for repeated or prolonged sweat monitoring throughout a day or week. Alternative methods must therefore be developed.

Fortunately, the remaining challenges for sweat biomonitoring do not seem to be fundamental impediments. As Gao and colleagues' work, and that of others^{2–6}, reveals the scale of the opportunities in this field, researchers will undoubtedly come up with innovations to transform technology that is currently merely appealing into something that, one day, you could not imagine living without. ■

Jason Heikenfeld is in the Department of Electrical Engineering and Computing Systems, University of Cincinnati, Cincinnati, Ohio 45221, USA.
e-mail: heikenjc@ucmail.uc.edu

1. Gao, W. *et al.* *Nature* **529**, 509–514 (2016).
2. Sonner, Z. *et al.* *Biomicrofluidics* **9**, 031301 (2015).
3. Bandothkar, A. J. & Wang, J. *Trends Biotechnol.* **32**, 363–371 (2014).
4. Huang, X. *et al.* *Small* **10**, 3083–3090 (2014).
5. Rose D. P. *et al.* *IEEE Trans. Biomed. Eng.* **62**, 1457–1465 (2015).
6. Matzeu, G. *et al.* *Anal. Meth.* **8**, 64–71 (2016).
7. www.idtechex.com/research/articles/idtechex-printed-electronics-usa-2014-award-winners-00007085.asp
8. www.eccrinesystems.com/sweatronicsplatform

The author declares competing financial interests: see go.nature.com/cwvyed for details.

Allowable CO₂ emissions based on regional and impact-related climate targets

Sonia I. Seneviratne¹, Markus G. Donat^{2,3}, Andy J. Pitman^{2,3}, Reto Knutti¹ & Robert L. Wilby⁴

Global temperature targets, such as the widely accepted limit of an increase above pre-industrial temperatures of two degrees Celsius, may fail to communicate the urgency of reducing carbon dioxide (CO₂) emissions. The translation of CO₂ emissions into regional- and impact-related climate targets could be more powerful because such targets are more directly aligned with individual national interests. We illustrate this approach using regional changes in extreme temperatures and precipitation. These scale robustly with global temperature across scenarios, and thus with cumulative CO₂ emissions. This is particularly relevant for changes in regional extreme temperatures on land, which are much greater than changes in the associated global mean.

The Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report included a figure in the Summary for Policymakers of Working Group 1 that linked global mean temperature changes (ΔT_{glob}) to total CO₂ emissions from 1870 onwards¹ (Fig. 1). This figure is compelling because it shows a clear linear relationship between cumulative CO₂ emissions and a measure of the global climate response. The obvious consequences are (1) that every tonne of CO₂ contributes about the same amount of global warming no matter when it is emitted, (2) that any target for the stabilization of ΔT_{glob} implies a finite CO₂ budget or quota that can be emitted, and (3) that global net emissions at some point need to be zero^{2–6}.

This simple relationship between CO₂ emissions and changes in ΔT_{glob} (Fig. 1) has helped overcome one communication barrier for the public in relating greenhouse gas emissions to the climate system response. However, there remains another obstacle to the full appreciation of associated climate impacts, namely, the translation of changes in global mean temperature to regional-scale consequences for society and the environment. In this Perspective, we demonstrate the feasibility and utility of quantitatively relating global cumulative CO₂ emissions to regional climate targets. We illustrate this approach by scaling changes in hot and cold extreme temperatures and heavy precipitation events with changes in the global mean temperature.

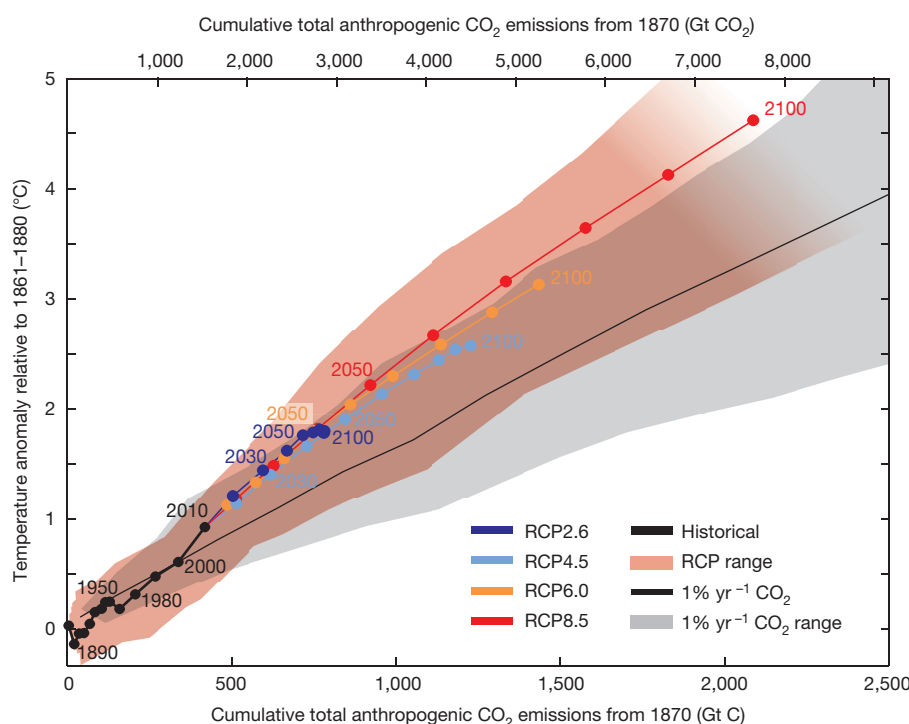


Figure 1 | Simulated global mean surface temperature increase as a function of cumulative total global CO₂ emissions. This is figure SPM.10 from ref. 1. It was derived from various lines of evidence. Model results over the historical period (1860–2010) are indicated in black. The coloured plume illustrates the multi-model spread over the four Representative Concentration Pathway (RCP) scenarios. The multi-model mean and range simulated by Coupled Model Intercomparison Project Phase 5 (CMIP5) models, forced by a CO₂ increase of 1% per year, is given by the thin black line and grey shading. For a given amount of cumulative CO₂ emissions, the 1% per year CO₂ simulations exhibit less warming than those driven by RCPs, which include additional non-CO₂ forcings. Temperature anomalies are given relative to the 1861–1880 base period; emissions are given relative to 1870.

¹Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, Switzerland. ²ARC Centre of Excellence in Climate System Science, University of New South Wales, Sydney, Australia. ³Climate Change Research Centre, University of New South Wales, Sydney, Australia. ⁴Department of Geography, Loughborough University, Loughborough, UK.

BOX 1

Linking regional extremes, global means, and cumulative emissions

We use output from the climate model simulations contributing to CMIP5⁵⁶. Here we present results for climate extreme indices representative of the hottest day (T_{Xx}) and coldest night (T_{Nn}) of the year, as well as the annual maximum consecutive 5-day precipitation total (R_{x5day}). Climate extreme indices⁵⁷ were calculated for the historical simulations⁵⁸ and future projections⁵⁹ from the CMIP5 ensemble. We use one run (r11p1) from models that provide historical simulations during 1861–2005, as well as RCP8.5 and RCP4.5 scenario simulations for the twenty-first century (see Supplementary Table 1). For the analysis of transient changes we concatenated historical (1861–2005) and RCP (2006–2099) simulations. We restricted our analyses to 1861–2099, which was common to all model runs. Global mean temperatures were calculated as the area-weighted global averages of annual mean temperatures. Extreme index fields were remapped to a common $2.5^\circ \times 2.5^\circ$ analysis grid to allow calculation of local ensemble averages and ensure that the same regions from each model contribute to the regional analyses.

Scatter plots showing the scaling relationship between changes in global mean temperature (ΔT_{glob}) and regional extreme index changes (see Figs 3 and 4b) are based on decadal averages of the respective variables. These averages of local anomalies relative to the 1861–1880 average were calculated for moving ten-year windows, and moving average values were assigned to the last year of each window period (that is, the value for year 2010 represents the average during 2001–2010; note that in the case of Fig. 1 the decadal global temperature averages are assigned to the year directly following that decade). These moving ten-year averages were also used to produce maps of local changes for a global mean temperature increase of 2°C (see Fig. 2). The indicated cumulative CO₂ emissions corresponding to different global mean temperature increases (red ticks on horizontal axes in Figs 3 and 4b) were approximated from the RCP8.5 ensemble average in Fig. 1 (single values were assigned to each of the chosen tick marks). This means that 500 billion tonnes of carbon (500 GtC) are emitted for a global increase of approximately 1.2°C , 1,000 GtC for 2.35°C , 1,500 GtC for 3.5°C , and 2,000 GtC for 4.45°C . Respective analyses regarding the scaling of extreme temperatures and precipitation in all 26 regions of the IPCC Special Report on Extremes⁷ and global land are provided in the Supplementary Information.

Global versus regional climate targets

Experience shows that the implications of projected global mean temperature changes tend to be underestimated at regional (and country) level, because the global changes are much smaller than the expected changes in regional temperature mean and extremes over most land areas^{7–10}. The limitations of global mean temperature as a measure of climate change have, for instance, been made evident by the public debate about the recent ‘hiatus’ decade in global warming, which has focused attention on changes in ΔT_{glob} instead of on the discernible worldwide impacts of the continued increases in radiative forcing^{1,11–14}.

As illustrated in Fig. 2, a 2°C target for ΔT_{glob} implies increases in both warm and cold temperature extremes that are greater than 2°C over most land regions. This is due to the land–sea contrast^{15,16} in response to radiative forcing, as well as to feedbacks (for example, from decreases in soil moisture, snow or ice^{7,8,17–20}), which further amplify changes in extreme temperatures in some key regions. As an example, the 2°C global mean

temperature target implies 3°C warming in hot temperature extremes in the Mediterranean region (Fig. 2a) and about 5.5°C warming in cold temperature extremes over land in the Arctic region (Fig. 2b). Hence, these changes in regional extremes are greater than those in global mean temperature by a factor of about 1.5 and 2.5–3 (Supplementary Fig. 1), respectively. As highlighted above, this stronger warming of extremes on land compared to that of global mean temperature is related both to the larger warming of mean temperature on land (Fig. 2c), as well as to an additional specific warming of extremes in several regions (Fig. 2a, b). Subjectively, such regional changes in extremes may convey the consequences of crossing the respective cumulative CO₂ emissions threshold better than the associated change in ΔT_{glob} (2°C), which seems relatively mild in comparison.

We make the case here for more easily interpretable analyses that relate global cumulative CO₂ emissions targets to changes in regional extremes or other impact-relevant quantities in addition to changes in global mean temperature. Although the IPCC Synthesis Report²¹ shows cumulative CO₂ emissions alongside their “reasons for concerns”, the bars (of various shades of red) provide only a qualitative assessment. We highlight here how quantitative analyses relating cumulative emissions to climate change at the national or regional scale could provide more targeted and actionable information for the decision process.

Regional extremes versus global CO₂ emissions

We thus assess the extent to which the implications of figure SPM.10 of ref. 1 (Fig. 1) can be expanded to relate cumulative global emissions in CO₂ with regional changes in temperature extremes (annual maximum and minimum temperatures; see Box 1). The result is displayed in Fig. 3 for four example regions with relatively strong scaling (the Mediterranean basin; the contiguous USA; central Brazil for annual maximum daytime temperatures; and the Arctic for annual minimum night-time temperatures). For other regions, see Supplementary Figs 4 and 5. The analyses display the scaling of the regional changes considered with the changes in global mean temperature for a range of climate projections, and provide the associated expected allowable cumulative global CO₂ emissions (but without considering the uncertainty in translating ΔT_{glob} to cumulative emissions).

The results show that changes in regional extreme temperatures display a rather linear scaling with ΔT_{glob} , which is also mostly independent of the emission scenario considered (Fig. 3). Hence, regional changes in temperature extremes can be usefully related to given cumulative CO₂ targets, without any consideration of the emission pathway. However, scaling for regional extremes on land is generally steeper than for ΔT_{glob} (see also analyses for other land regions in Supplementary Figs 4 and 5). Hence, as expected from Fig. 2, the relationship between the increase in regional temperature extremes and the increase in global mean temperature typically implies a larger change of the former at more local scales.

For instance, a 2°C warming in hot extremes (annual warmest daytime temperature, T_{Xx}) takes place in the Mediterranean for a change of 1.4°C in ΔT_{glob} (Fig. 3a). The corresponding allowable cumulative CO₂ emissions are therefore about 600 Gt C for a 2°C warming of hot extremes in the Mediterranean region compared to about 850 Gt C for a 2°C warming in global mean. Given current political tensions around the Mediterranean basin, implications of locally more rapid climate change could extend to regional impacts²², adding to wider political instability (see for example the purported impacts of drought in Syria^{23,24}).

Scaling extreme hot temperatures in the contiguous USA and central Brazil (Fig. 3b, c) by ΔT_{glob} provides qualitatively similar results, but highlights greater uncertainty of projections in these regions. In the contiguous USA, although the expected value of scaling with ΔT_{glob} is greater than 1, the uncertainty range bounds the 1:1 (identity) line. Conversely, the regional response in central Brazil is significantly different from the 1:1 line despite the larger uncertainty range compared to the Mediterranean region. The response of the regional changes in annual coldest night-time temperatures (T_{Nn}) in the Arctic (Fig. 3d) conveys a very stark message. In this case, as seen in Fig. 2, the regional response is about 2.5–3 times greater for

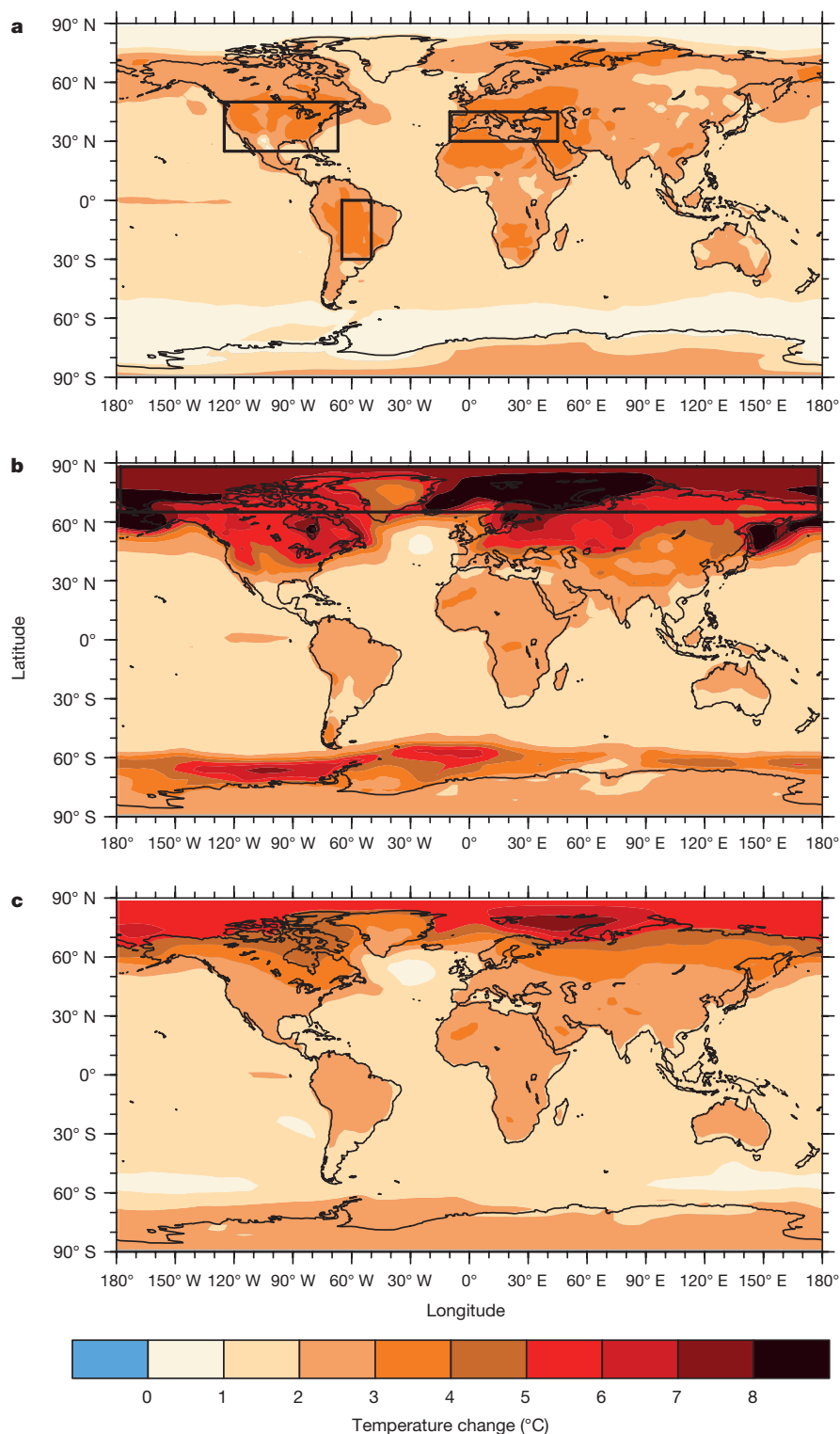


Figure 2 | Extreme (and mean) temperature changes associated with a 2°C target. Local changes associated with a global warming of 2°C are shown for hottest daytime temperature (T_{Xx}) (a), annual coldest night-time temperature (T_{Nn}) (b), and mean temperature (T_{mean}) (c). The analysis is based on RCP8.5 scenario simulations (ensemble average year 2044; based on 25 model simulations, see Supplementary Table 1).

The respective scaling expressed as ratio of global mean temperature increase is provided in Supplementary Fig. 1. Note that very similar results are obtained with the RCP4.5 scenario simulations (Fig. 3 and Supplementary Figs 2 and 3, based on 22 model simulations). Panels a and b also display the outlines of the regions analysed in Fig. 3.

the coldest extremes than for the global mean temperature change, with an increase of about 5.5°C for the 2°C global warming target. In addition, it is evident that a regional 2°C threshold was passed in the simulations around the year 2000 for T_{Nn} in the Arctic, while it is projected to be reached by about 2030 for T_{Xx} in the Mediterranean, central Brazil and the contiguous

USA, and only by the mid-2040s for the global mean temperature, under the business-as-usual (unchecked) emissions scenario (RCP8.5, which leads to a radiative forcing of 8.5 W m^{-2} by 2100 relative to pre-industrial values). For a 1.5°C global warming target, we also note that substantial regional changes in temperature extremes would still occur, with (for example) a

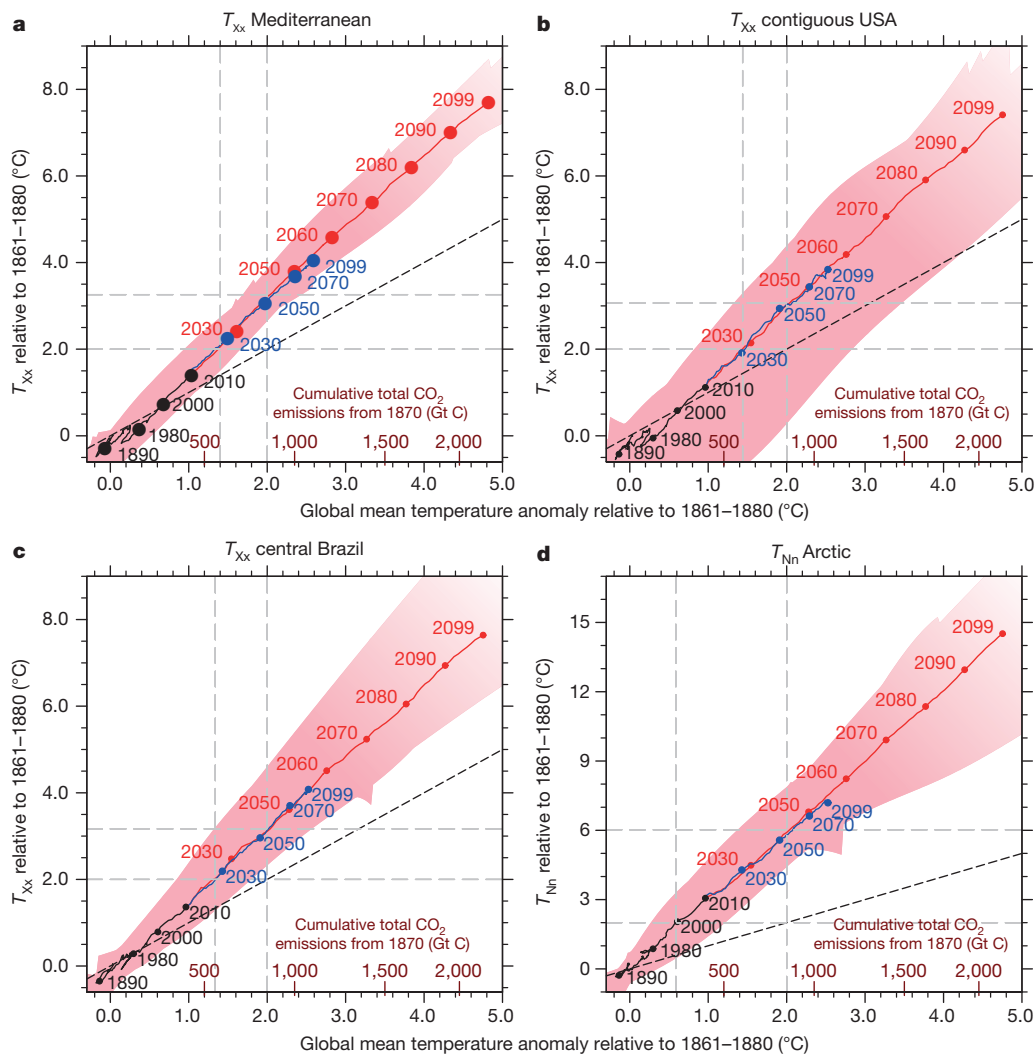


Figure 3 | Scaling between regional changes in annual temperature extremes and changes in global mean temperature, with associated global cumulative CO₂ emissions targets. See Box 1 for details on the underlying analysis. Results are shown for annual maximum daytime temperature (T_{xx}) in the Mediterranean region (30° to 45° N, 10° W to 45° E) (a), the contiguous USA (25° to 50° N, 125° W to 67° W) (b), and central Brazil (30° S to 0° N, 65° W to 50° W) (c), and for the annual minimum night-time temperature (T_{nn}) in the Arctic (65° to 90° N, 180° W to 180° E) (d). The four analysed regions are indicated in Fig. 2a and b. The solid black line denotes the ensemble average in the historical runs

4.4°C warming in T_{nn} in the Arctic and a 2.2°C increase in T_{xx} in the Mediterranean region (Fig. 3).

Although we have illustrated the concept of regional and impact-related climate targets with regional changes in temperature extremes, similar reasoning can be applied to a range of other responses to global climate forcing^{7,25} (for example, changes in heavy precipitation events, see below). These are also highly relevant in comprehending the regional implications of global CO₂ emissions. As a further illustration, we display in Fig. 4 the scaling of heavy precipitation events with global mean temperature, and the respective relationship between cumulative CO₂ emissions and resulting changes in heavy precipitation in Southern Asia. As for regional temperature extremes, multi-model average changes in heavy precipitation display an almost linear scaling with the changes in global mean temperature²⁶ (roughly consistent with the Clausius–Clapeyron relationship in that region), and thus could be used to provide regional decision-makers with suitable allowable targets for global emissions.

Moreover, it should be noted that, while the ensemble mean response is robust across models and emissions scenarios for heavy precipitation events, individual model projections can diverge strongly from this

mean response (in the region investigated as well as in other locations; see Supplementary Figs 6 and 7). This is obvious from the red-shaded uncertainty range in Fig. 4b and Supplementary Figs 6 and 7, which is substantially larger in most regions than for temperature extremes. This behaviour is due to the increasing relevance of internal climate variability at the regional-to-local scale²⁷, higher model uncertainty, and the spatially more heterogeneous nature of precipitation extremes compared to temperature extremes.

Despite the associated uncertainty, analyses such as the ones in Figs 3 and 4b provide more information to regional stakeholders than a global mean temperature target, since they quantitatively and directly highlight the expected regional response (in extremes and other variables than temperature), with attendant lower and upper bounds. Such estimates are thus more useful when assessing associated impacts, and engaging with policymakers.

Limitations of approach

Some caveats are attached to the above findings, most importantly:

(1) Scaling relationships are only meaningful as long as associated uncertainties in projections are kept within reasonable bounds. This

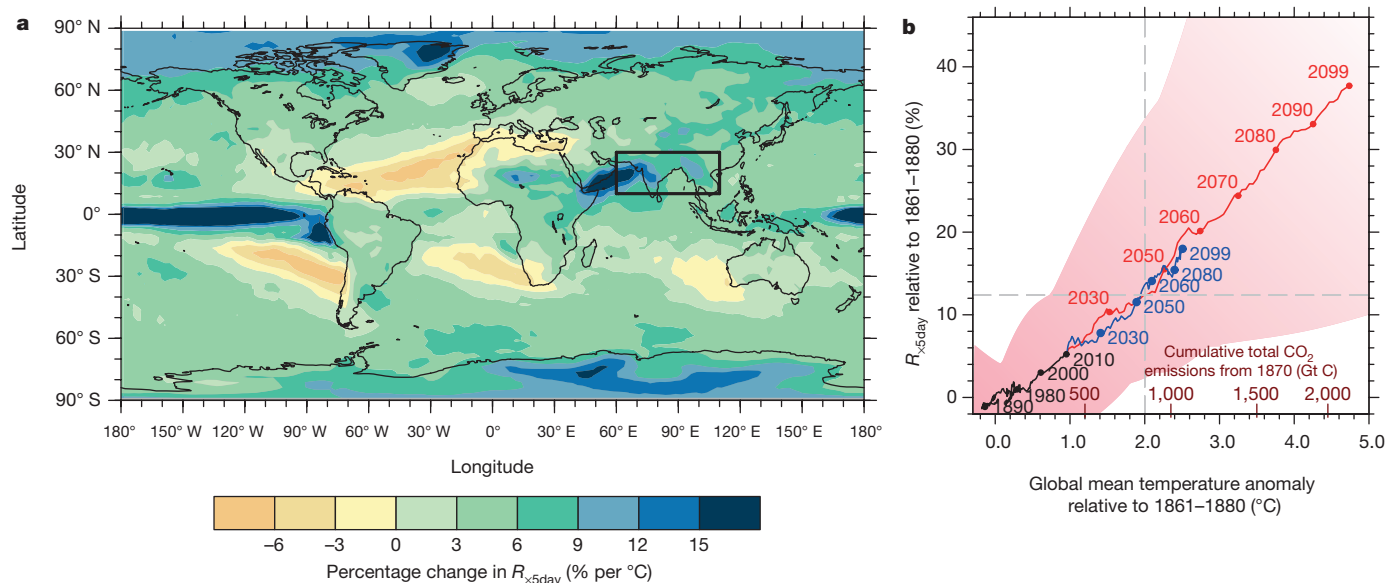


Figure 4 | Scaling of 5-day heavy precipitation events with global mean temperature changes, with associated global cumulative CO₂ emissions targets. See Box 1 for details on the underlying analysis. **a**, Map of the ratio of percentage changes in heavy precipitation events (annual maximum consecutive 5-day precipitation, R_{x5day}) with changes in global mean temperature ΔT_{glob} for the RCP8.5 scenario simulations (ensemble average ratio $\Delta R_{x5day}/\Delta T_{glob}$). ΔT_{glob} and ΔR_{x5day} were calculated from each model run as the difference between the average of the first (1861–1880) and last (2080–2099) 20-year time slices. **b**, Scaling of percentage changes in R_{x5day} in Southern Asia (10° to 30° N, 60° to 110° E; see outlined box in **a**) with global mean temperature changes

is the case for some climate features, such as temperature extremes or heavy precipitation events^{1,7}, but for others, such as droughts, tropical cyclones or storms, uncertainties are generally larger than the climate change signals^{1,7,28}. In such situations, no emissions target (or implied global temperature target) may currently be set on the basis of avoiding changes in these extremes.

(2) Some changes in the climate system may be abrupt (that is, nonlinearly related to emissions) owing to tipping points²⁹. Again, uncertainties in the associated projections are very large, especially under high-end emissions. Owing to the nonlinearity of the respective features, relationships could be difficult to derive (although some features have been assessed, such as the dependency of mean sea level rise on global mean temperature increase at equilibrium³⁰ and the probability of abrupt changes for given global temperature thresholds³¹).

(3) Although we find a relatively robust scaling of regional temperature and precipitation extremes with ΔT_{glob} , we can expect that the reliability of scaling will diminish at increasingly smaller scales owing to internal climate variability^{27,32} and a larger contribution of local processes to the response (including by local land surface and human forcing, see point (5) below).

(4) It is likely that climate models share common biases for some regional climate phenomena^{33–36}. In this case, scaling features could be derived, but would be erroneous, an issue that would need to be examined with careful model evaluation^{37,38} contingent on the availability of appropriate observations.

(5) The relationship between changes in regional climate and ΔT_{glob} would be expected to alter in the presence of time-varying local forcing by, for example, aerosols³⁹, land-use and land-cover change^{40–42}, urban development⁴³, or human water use^{44,45}. These effects are likely to be important on the local scale, but less so for the larger regions considered here (see Figs 3 and 4 and the regions from the IPCC Special Report on Extremes⁷ in the Supplementary Information).

(6) The ranges in Figs 3 and 4b reflect the uncertainty in the scaling of the regional quantities with ΔT_{glob} , but do not include uncertainties

and cumulative global CO₂ emissions. The solid black line denotes the ensemble average in the historical runs until 2010 (combined with RCP8.5 for 2006–2010), and the solid red (blue) line denotes the ensemble average of the future projections following the RCP8.5 (RCP4.5) scenario simulations, based on 25 (22) model simulations (see Supplementary Table 1). The red shaded area indicates the total range (minimum to maximum value) for all considered simulations and experiments. Grey dashed lines show the percentage change in R_{x5day} or CO₂ emissions associated with a 2°C increase in global mean temperature. Only land grid cells were used for calculating the regional R_{x5day} average.

associated with the scaling of ΔT_{glob} with the cumulative CO₂ emissions (Fig. 1). This additional uncertainty source is also relevant for the decision process when assessing regional climate targets (as is the case for climate targets based on the global mean temperature). For a given impact threshold, the uncertainty in the cumulative carbon would be wider, and as a consequence the cumulative carbon budget would be smaller if the desire were to avoid the impact with high probability⁵. More in-depth analyses of the CMIP5 archive would help determine the total uncertainty range when directly relating imposed greenhouse gas forcing to simulated regional extremes.

Using regional targets in decision making

We focus here on regional changes because local stakeholders and decision-makers are more likely to be able to relate to them than to global mean temperature changes. However, we stress that this does not imply that countries should only be concerned about climate changes affecting them directly in a geographical sense. Indeed, because of globalization, major climate disruptions in some countries can strongly affect others, for instance owing to political unrest, migration, impacts on global food production, supply chains and trade^{23,46,47}. Even when not directly affected by such changes, individual countries are more likely to understand the implications of climate targets for other parties if they can more readily quantify the specific implications for different regions. This could also help pave the way to solutions that integrate both climate mitigation and adaptation within climate negotiations, by incorporating the costs of impacts into negotiations. Global temperature targets that differ from and are possibly lower than 2°C (such as 1.5°C)^{48–50} may thus well be desirable on the basis of inferred regional climate targets.

Linking cumulative CO₂ emission targets to regional consequences, such as changing climate extremes, would be of particular benefit for political decision-making, both in the context of climate negotiations and of adaptation. We stress that the quantification of regional targets will not necessarily imply that all involved parties will agree on a suitable (and common) cumulative global CO₂ emission target. However,

regional information can help in the development of solutions and in communication with the public. Similarly robust regional scaling might be expected for other features of the climate system beside those considered here^{51,52}, and could be explored for impact-based simulations^{53–55}. Indeed, such relationships can be determined for any regional or impact-relevant climate feature that scales robustly with changes in global mean temperature (or is at least monotonically related to it), and that is not associated with larger uncertainty ranges or biases in current climate models.

In view of the inherent model uncertainty and to avoid possible risks associated with the indiscriminate use of such information, we recommend that IPCC-calibrated language be applied when assessing the confidence of any such derived relationships, with only situations of 'high confidence' justifying the derivation of quantitative estimates⁷. In addition to the requirement of levels of high confidence, a high signal-to-noise model ratio (traditionally referred to in 'likelihood' terms in the language⁷ of the IPCC) is a prerequisite for deriving meaningful allowable CO₂ emissions ranges. Furthermore, any assessment of projected changes in climate risks and impacts also needs to consider the contributions of changes in vulnerability and exposure of human and natural systems to those climate hazards²⁵. Bearing in mind these requirements, quantitative tools for decision-making that relate regional (or even country-scale) impacts to global CO₂ emissions targets could be one way of advancing climate negotiations by exposing what is at stake in a more local manner.

Received 17 September; accepted 16 December 2015.

Published online 20 January 2016.

- Intergovernmental Panel on Climate Change (IPCC). In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 3–29 (Cambridge Univ. Press, 2013). **This 'Summary for Policymakers' (approved line by line by the IPCC plenary) includes, for the first time, a figure relating cumulative CO₂ emissions with projected changes in global mean temperature (Fig. 1 in this Perspective); it builds upon refs 2–4 and more recent simulations and publications on this topic.**
- Meinshausen, M. et al. Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature* **458**, 1158–1162 (2009).
- Allen, M. R. et al. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).
- Matthews, H. D., Gillett, N. P., Stott, P. A. & Zickfeld, K. The proportionality of global warming to cumulative carbon emissions. *Nature* **459**, 829–832 (2009).
- Knutti, R. & Rogelj, J. The legacy of our CO₂ emissions: a clash of scientific facts, politics and ethics. *Clim. Change* **133**, 361–373 (2015).
- Friedlingstein, P. et al. Persistent growth of CO₂ emissions and implications for reaching climate targets. *Nature Geosci.* **7**, 709–715 (2014).
- Seneviratne, S. I. et al. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change 109–230 (Cambridge Univ. Press, 2012).
- Orlowsky, B. & Seneviratne, S. I. Global changes in extreme events: regional and seasonal dimension. *Clim. Change* **110**, 669–696 (2012). **This article provides an analysis of the scaling of changes in regional temperature extremes with changes in global warming, as well as its decomposition in several contributing factors (regional, seasonal, and differential response of extremes versus the median).**
- Lehner, F. & Stocker, T. F. From local perception to global perspective. *Nature Clim. Change* **5**, 731–734 (2015).
- Diffenbaugh, N. S. & Ashfaq, M. Intensification of hot extremes in the United States. *Geophys. Res. Lett.* **37**, L15701 (2010).
- Trenberth, K. E. & Fasullo, J. T. An apparent hiatus in global warming? *Earth's Future* **1**, 19–32 (2013).
- Seneviratne, S. I., Donat, M., Mueller, B. & Alexander, L. V. No pause in the increase of hot temperature extremes. *Nature Clim. Change* **4**, 161–163 (2014).
- Victor, D. G. & Kennel, C. F. Climate policy: Ditch the 2 °C warming goal. *Nature* **514**, 30–31 (2014).
- Karl, T. R. et al. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**, 1469–1472 (2015).
- Sutton, R. T., Dong, B. & Gregory, J. M. Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations. *Geophys. Res. Lett.* **34**, L02701 (2007).
- Herger, N., Sanderson, B. M. & Knutti, R. Improved pattern scaling approaches for the use in climate impact studies. *Geophys. Res. Lett.* **42**, 3486–3494 (2015).
- Seneviratne, S. I., Lüthi, D., Litschi, M. & Schär, C. Land-atmosphere coupling and climate change in Europe. *Nature* **443**, 205–209 (2006).
- Kharin, V. V., Zwiers, F. W., Zhang, X. & Hegerl, G. C. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* **20**, 1419–1444 (2007).
- Seneviratne, S. I. et al. Impact of soil moisture-climate feedbacks on CMIP5 projections: first results from the GLACE-CMIP5 experiment. *Geophys. Res. Lett.* **40**, 5212–5217 (2013).
- Serreze, M. C. & Barry, R. G. Processes and impacts of Arctic amplification. A research synthesis. *Glob. Planet. Change* **77**, 85–96 (2011).
- Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds The Core Writing Team, Pachauri, R. K. & Meyer, L. A.) 1–151 (IPCC, 2014).
- Council of the European Union. Climate Change and International Security <http://register.consilium.europa.eu/doc/srv?l=EN&f=ST%207249%202008%20INIT> [accessed 16 November 2015] (2008).
- Kelley, C. P., Mohtadi, S., Cane, M. A., Seager, R. & Kushnir, Y. Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proc. Natl Acad. Sci. USA* **112**, 3241–3246 (2015).
- Murray, V. et al. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. et al.) A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC) 487–542 (Cambridge Univ. Press, 2012).
- Intergovernmental Panel on Climate Change (IPCC). In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C. B. et al.) 1–32 (Cambridge Univ. Press, 2014).
- Fischer, E. M., Sedláček, J., Hawkins, E. & Knutti, R. Models agree on forced response pattern of precipitation and temperature extremes. *Geophys. Res. Lett.* **41**, 8554–8562 (2014). **This article shows a substantial intermodel agreement of the forced response pattern of precipitation and temperature extremes.**
- Deser, C., Knutti, R., Solomon, S. & Phillips, A. Communication of the role of natural variability in future North American climate. *Nature Clim. Change* **2**, 775–779 (2012).
- Orlowsky, B. & Seneviratne, S. I. Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections. *Hydrol. Earth Syst. Sci.* **17**, 1765–1781 (2013).
- Lenton, T. M. et al. Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
- Church, J. A. et al. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1137–1216 (Cambridge Univ. Press, 2013).
- Drijfhout, S. et al. Catalogue of abrupt shifts in Intergovernmental Panel on Climate Change climate models. *Proc. Natl. Acad. Sci.* **112**, E5777–E5786 (2015).
- Sutton, R., Suckling, E. & Hawkins, E. What does global mean temperature tell us about local climate? *Phil. Trans. R. Soc. A* **373**, 20140426 (2015).
- Flato, G. et al. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 741–866 (Cambridge Univ. Press, 2013).
- Taylor, C. M., de Jeu, R. A. M., Guichard, F., Harris, P. P. & Dorigo, W. A. Afternoon rain more likely over drier soils. *Nature* **489**, 423–426 (2012).
- Mueller, B. & Seneviratne, S. I. Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.* **41**, 128–134 (2014).
- Masato, G., Hoskins, B. & Woollings, T. Winter and summer Northern Hemisphere blocking in CMIP5 models. *J. Clim.* **26**, 7044–7059 (2013).
- Hall, A. & Qu, X. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.* **33**, L03502 (2006).
- Boisier, J. P., Ciais, P., Ducharme, A. & Guimberteau, M. Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nature Clim. Change* **5**, 656–660 (2015).
- Levy, H. II et al. The role of aerosol direct and indirect effects in past and future climate change. *J. Geophys. Res.* **118**, 4521–4532 (2013).
- Pitman, A. J. et al. Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study. *Geophys. Res. Lett.* **36**, L14814 (2009).
- Luyssaert, S. et al. Land management and land-cover changes have impacts of similar magnitude on surface temperature. *Nature Clim. Change* **4**, 389–393 (2014).
- Jeong, S.-J. et al. Effects of double cropping on summer climate of the North China Plain and neighbouring regions. *Nature Clim. Change* **4**, 615–619 (2014).
- Wilby, R. L. Constructing climate change scenarios of urban heat island intensity and air quality. *Environ. Plann. B* **35**, 902–919 (2008).
- Wei, J., Dirmeyer, P. A., Wiser, D., Bosilovich, M. G. & Mocko, D. M. Where does the irrigation water go? An estimate of the contribution of irrigation to precipitation using MERRA. *J. Hydrometeorol.* **14**, 275–289 (2013).
- Degu, A. H. et al. The influence of large dams on surrounding climate and precipitation patterns. *Geophys. Res. Lett.* **38**, L04405 (2011).
- Orlowsky, B., Hoekstra, A. Y., Gudmundsson, L. & Seneviratne, S. I. Today's virtual water consumption and trade under future water scarcity. *Environ. Res. Lett.* **9**, 074007 (2014).

47. Hunt, A. S. P., Wilby, R. L., Dale, N., Sura, K. & Watkiss, P. Embodied water imports to the UK under climate change. *Clim. Res.* **59**, 89–101 (2014).
 48. Hansen, J. *et al.* Assessing “dangerous climate change”: required reduction of carbon emissions to protect young people, future generations, and nature. *PLoS ONE* **8**, e81648 (2013).
 49. Tschakert, P. 1.5°C or 2°C: a conduit’s view from the science-policy interface at COP20 in Lima, Peru. *Clim. Change Resp.* **2**, 3 (2015).
 50. United Nations Framework Convention on Climate Change (UNFCCC). *Report on the Structured Expert Dialogue on the 2013–2015 Review (FCCC/SB/2015/INF.1)* 1–182, <http://unfccc.int/resource/docs/2015/sb/eng/inf01.pdf> [accessed 16 November 2015] (UNFCCC, 2015).
 51. Christensen, J. H. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 1217–1308 (Cambridge Univ. Press, 2013).
 52. Frieler, K., Meinshausen, M., Mengel, M., Braun, N. & Hare, W. A scaling approach to probabilistic assessment of regional climate change. *J. Clim.* **25**, 3117–3144 (2012).
 53. Schewe, J. *et al.* Multimodel assessment of water scarcity under climate change. *Proc. Natl Acad. Sci. USA* **111**, 3245–3250 (2014).
 54. Hanewinkel, M., Cullmann, D. A., Schelhaas, M.-J., Nabuurs, G.-J. & Zimmermann, N. E. Climate change may cause severe loss in the economic value of European forest land. *Nature Clim. Change* **3**, 203–207 (2012).
 55. Pal, J. S. & Eltahir, E. A. B. Future temperature in southwest Asia projected to exceed a threshold for human adaptability. *Nature Clim. Change* <http://dx.doi.org/doi:10.1038/nclimate2833> (2015).
 56. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
 57. Zhang, X. *et al.* Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisc. Rev. Clim. Change* **2**, 851–870 (2011).
 58. Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.* **118**, 1716–1733 (2013).
 59. Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: part 2. Future climate projections. *J. Geophys. Res. Atmos.* **118**, 2473–2493 (2013).
- This article provides time series of climate extreme indices in CMIP5 projections, which have been used as the basis for the present analyses.**

Supplementary Information is available in the online version of the paper.

Acknowledgements S.I.S. acknowledges the European Research Council (ERC) ‘DROUGHT-HEAT’ project funded by the European Community’s Seventh Framework Programme (grant agreement FP7-IDEAS-ERC-617518). A.J.P. and M.G.D. were supported by the Australian Research Council (ARC) Centre of Excellence for Climate System Science (grant number CE110001028). M.G.D. was also supported by the ARC (grant number DE150100456). This work contributes to the World Climate Research Programme (WCRP) Grand Challenge on Extremes. We acknowledge the WCRP Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank N. Maher for help with processing CMIP5 data. The climate extremes indices calculated for the different CMIP5 runs were obtained from the Environment Canada CLIMDEX website (<http://www.cccma.ec.gc.ca/data/climdex/>).

Author Contributions S.I.S., M.G.D. and A.J.P. designed the study, following an initial discussion between S.I.S., A.J.P. and R.K. S.I.S. coordinated the conception and writing of the article. M.G.D. performed the analyses. R.L.W. contributed to the interpretation of regional impacts. All authors commented on the manuscript and analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.I.S. (sonia.seneviratne@ethz.ch).

Mastering the game of Go with deep neural networks and tree search

David Silver^{1*}, Aja Huang^{1*}, Chris J. Maddison¹, Arthur Guez¹, Laurent Sifre¹, George van den Driessche¹, Julian Schrittwieser¹, Ioannis Antonoglou¹, Veda Panneershelvam¹, Marc Lanctot¹, Sander Dieleman¹, Dominik Grewe¹, John Nham², Nal Kalchbrenner¹, Ilya Sutskever², Timothy Lillicrap¹, Madeleine Leach¹, Koray Kavukcuoglu¹, Thore Graepel¹ & Demis Hassabis¹

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses ‘value networks’ to evaluate board positions and ‘policy networks’ to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm, our program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0. This is the first time that a computer program has defeated a human professional player in the full-sized game of Go, a feat previously thought to be at least a decade away.

All games of perfect information have an optimal value function, $v^*(s)$, which determines the outcome of the game, from every board position or state s , under perfect play by all players. These games may be solved by recursively computing the optimal value function in a search tree containing approximately b^d possible sequences of moves, where b is the game’s breadth (number of legal moves per position) and d is its depth (game length). In large games, such as chess ($b \approx 35$, $d \approx 80$)¹ and especially Go ($b \approx 250$, $d \approx 150$)¹, exhaustive search is infeasible^{2,3}, but the effective search space can be reduced by two general principles. First, the depth of the search may be reduced by position evaluation: truncating the search tree at state s and replacing the subtree below s by an approximate value function $v(s) \approx v^*(s)$ that predicts the outcome from state s . This approach has led to superhuman performance in chess⁴, checkers⁵ and othello⁶, but it was believed to be intractable in Go due to the complexity of the game⁷. Second, the breadth of the search may be reduced by sampling actions from a policy $p(a|s)$ that is a probability distribution over possible moves a in position s . For example, Monte Carlo rollouts⁸ search to maximum depth without branching at all, by sampling long sequences of actions for both players from a policy p . Averaging over such rollouts can provide an effective position evaluation, achieving superhuman performance in backgammon⁸ and Scrabble⁹, and weak amateur level play in Go¹⁰.

Monte Carlo tree search (MCTS)^{11,12} uses Monte Carlo rollouts to estimate the value of each state in a search tree. As more simulations are executed, the search tree grows larger and the relevant values become more accurate. The policy used to select actions during search is also improved over time, by selecting children with higher values. Asymptotically, this policy converges to optimal play, and the evaluations converge to the optimal value function¹². The strongest current Go programs are based on MCTS, enhanced by policies that are trained to predict human expert moves¹³. These policies are used to narrow the search to a beam of high-probability actions, and to sample actions during rollouts. This approach has achieved strong amateur play^{13–15}. However, prior work has been limited to shallow

policies^{13–15} or value functions¹⁶ based on a linear combination of input features.

Recently, deep convolutional neural networks have achieved unprecedented performance in visual domains: for example, image classification¹⁷, face recognition¹⁸, and playing Atari games¹⁹. They use many layers of neurons, each arranged in overlapping tiles, to construct increasingly abstract, localized representations of an image²⁰. We employ a similar architecture for the game of Go. We pass in the board position as a 19×19 image and use convolutional layers to construct a representation of the position. We use these neural networks to reduce the effective depth and breadth of the search tree: evaluating positions using a value network, and sampling actions using a policy network.

We train the neural networks using a pipeline consisting of several stages of machine learning (Fig. 1). We begin by training a supervised learning (SL) policy network p_σ directly from expert human moves. This provides fast, efficient learning updates with immediate feedback and high-quality gradients. Similar to prior work^{13,15}, we also train a fast policy p_π that can rapidly sample actions during rollouts. Next, we train a reinforcement learning (RL) policy network p_ρ that improves the SL policy network by optimizing the final outcome of games of self-play. This adjusts the policy towards the correct goal of winning games, rather than maximizing predictive accuracy. Finally, we train a value network v_θ that predicts the winner of games played by the RL policy network against itself. Our program AlphaGo efficiently combines the policy and value networks with MCTS.

Supervised learning of policy networks

For the first stage of the training pipeline, we build on prior work on predicting expert moves in the game of Go using supervised learning^{13,21–24}. The SL policy network $p_\sigma(a|s)$ alternates between convolutional layers with weights σ , and rectifier nonlinearities. A final softmax layer outputs a probability distribution over all legal moves a . The input s to the policy network is a simple representation of the board state (see Extended Data Table 2). The policy network is trained on randomly

¹Google DeepMind, 5 New Street Square, London EC4A 3TW, UK. ²Google, 1600 Amphitheatre Parkway, Mountain View, California 94043, USA.

*These authors contributed equally to this work.

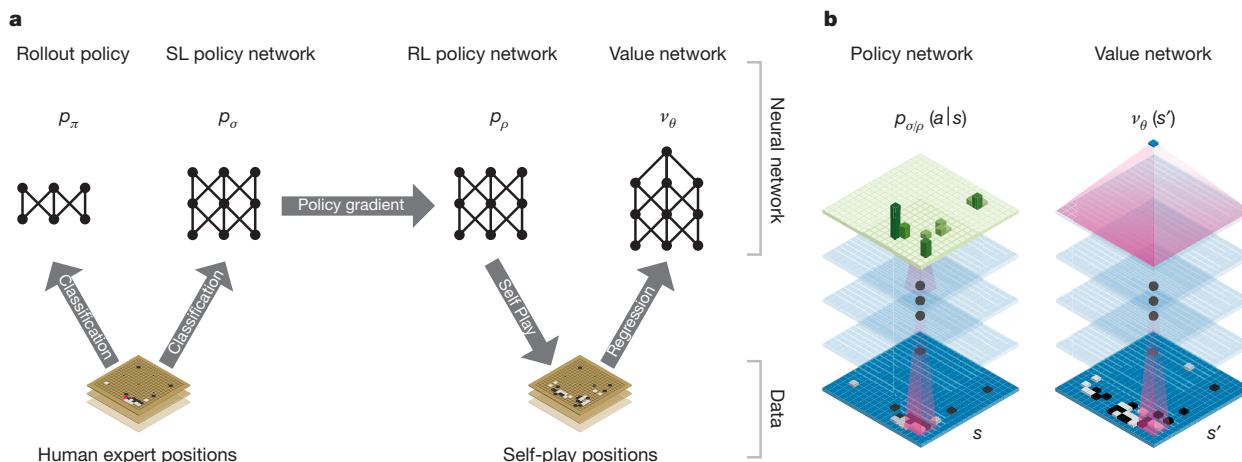


Figure 1 | Neural network training pipeline and architecture. **a**, A fast rollout policy p_π and supervised learning (SL) policy network p_σ are trained to predict human expert moves in a data set of positions. A reinforcement learning (RL) policy network p_ρ is initialized to the SL policy network, and is then improved by policy gradient learning to maximize the outcome (that is, winning more games) against previous versions of the policy network. A new data set is generated by playing games of self-play with the RL policy network. Finally, a value network v_θ is trained by regression to predict the expected outcome (that is, whether

the current player wins) in positions from the self-play data set.

b, Schematic representation of the neural network architecture used in AlphaGo. The policy network takes a representation of the board position s as its input, passes it through many convolutional layers with parameters σ (SL policy network) or ρ (RL policy network), and outputs a probability distribution $p_\sigma(a|s)$ or $p_\rho(a|s)$ over legal moves a , represented by a probability map over the board. The value network similarly uses many convolutional layers with parameters θ , but outputs a scalar value $v_\theta(s')$ that predicts the expected outcome in position s' .

sampled state-action pairs (s, a) , using stochastic gradient ascent to maximize the likelihood of the human move a selected in state s

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial \sigma}$$

We trained a 13-layer policy network, which we call the SL policy network, from 30 million positions from the KGS Go Server. The network predicted expert moves on a held out test set with an accuracy of 57.0% using all input features, and 55.7% using only raw board position and move history as inputs, compared to the state-of-the-art from other research groups of 44.4% at date of submission²⁴ (full results in Extended Data Table 3). Small improvements in accuracy led to large improvements in playing strength (Fig. 2a); larger networks achieve better accuracy but are slower to evaluate during search. We also trained a faster but less accurate rollout policy $p_\pi(a|s)$, using a linear softmax of small pattern features (see Extended Data Table 4) with weights π ; this achieved an accuracy of 24.2%, using just 2 μ s to select an action, rather than 3 ms for the policy network.

Reinforcement learning of policy networks

The second stage of the training pipeline aims at improving the policy network by policy gradient reinforcement learning (RL)^{25,26}. The RL policy network p_ρ is identical in structure to the SL policy network,

and its weights ρ are initialized to the same values, $\rho = \sigma$. We play games between the current policy network p_ρ and a randomly selected previous iteration of the policy network. Randomizing from a pool of opponents in this way stabilizes training by preventing overfitting to the current policy. We use a reward function $r(s)$ that is zero for all non-terminal time steps $t < T$. The outcome $z_t = \pm r(s_T)$ is the terminal reward at the end of the game from the perspective of the current player at time step t : +1 for winning and -1 for losing. Weights are then updated at each time step t by stochastic gradient ascent in the direction that maximizes expected outcome²⁵

$$\Delta\rho \propto \frac{\partial \log p_\rho(a_t|s_t)}{\partial \rho} z_t$$

We evaluated the performance of the RL policy network in game play, sampling each move $a_t \sim p_\rho(\cdot|s_t)$ from its output probability distribution over actions. When played head-to-head, the RL policy network won more than 80% of games against the SL policy network. We also tested against the strongest open-source Go program, Pachi¹⁴, a sophisticated Monte Carlo search program, ranked at 2 amateur *dan* on KGS, that executes 100,000 simulations per move. Using no search at all, the RL policy network won 85% of games against Pachi. In comparison, the previous state-of-the-art, based only on supervised

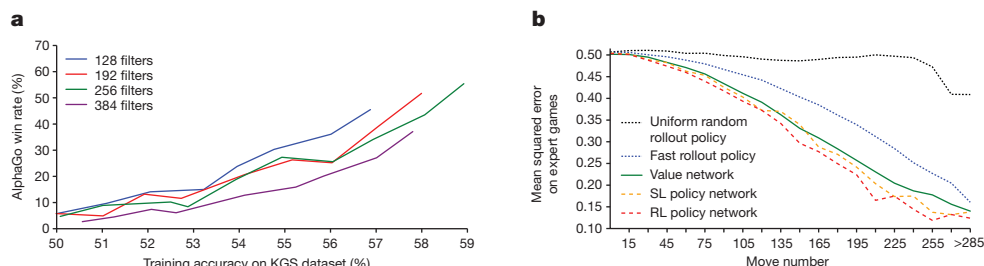


Figure 2 | Strength and accuracy of policy and value networks.

a, Plot showing the playing strength of policy networks as a function of their training accuracy. Policy networks with 128, 192, 256 and 384 convolutional filters per layer were evaluated periodically during training; the plot shows the winning rate of AlphaGo using that policy network against the match version of AlphaGo. **b**, Comparison of evaluation accuracy between the value network and rollouts with different policies.

Positions and outcomes were sampled from human expert games. Each position was evaluated by a single forward pass of the value network v_θ , or by the mean outcome of 100 rollouts, played out using either uniform random rollouts, the fast rollout policy p_π , the SL policy network p_σ or the RL policy network p_ρ . The mean squared error between the predicted value and the actual game outcome is plotted against the stage of the game (how many moves had been played in the given position).

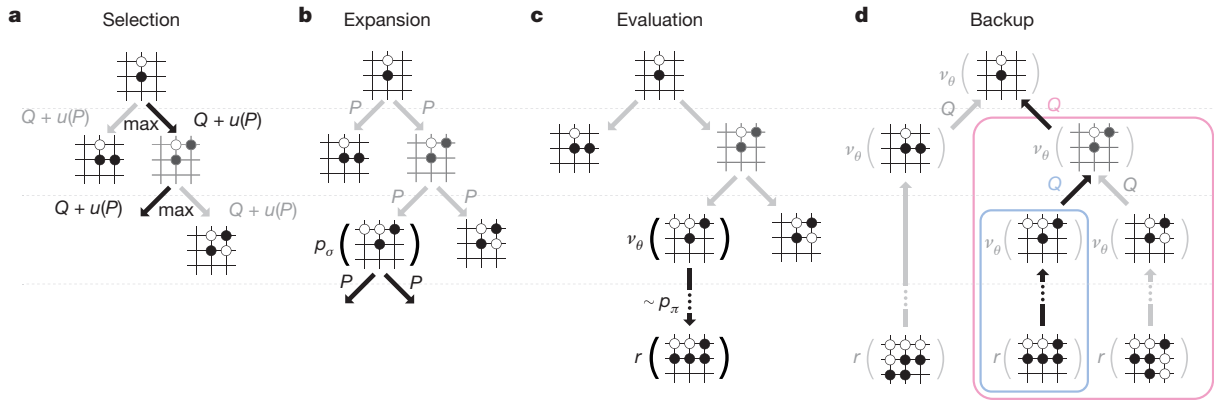


Figure 3 | Monte Carlo tree search in AlphaGo. **a**, Each simulation traverses the tree by selecting the edge with maximum action value Q , plus a bonus $u(P)$ that depends on a stored prior probability P for that edge. **b**, The leaf node may be expanded; the new node is processed once by the policy network p_σ and the output probabilities are stored as prior probabilities P for each action. **c**, At the end of a simulation, the leaf node

learning of convolutional networks, won 11% of games against Pachi²³ and 12% against a slightly weaker program, Fuego²⁴.

Reinforcement learning of value networks

The final stage of the training pipeline focuses on position evaluation, estimating a value function $v^p(s)$ that predicts the outcome from position s of games played by using policy p for both players^{28–30}

$$v^p(s) = \mathbb{E}[z_t | s_t = s, a_{1:T} \sim p]$$

Ideally, we would like to know the optimal value function under perfect play $v^*(s)$; in practice, we instead estimate the value function v_{p_σ} for our strongest policy, using the RL policy network p_σ . We approximate the value function using a value network $v_\theta(s)$ with weights θ , $v_\theta(s) \approx v_{p_\sigma}(s) \approx v^*(s)$. This neural network has a similar architecture to the policy network, but outputs a single prediction instead of a probability distribution. We train the weights of the value network by regression on state–outcome pairs (s, z) , using stochastic gradient descent to minimize the mean squared error (MSE) between the predicted value $v_\theta(s)$, and the corresponding outcome z

$$\Delta\theta \propto \frac{\partial v_\theta(s)}{\partial \theta} (z - v_\theta(s))$$

The naive approach of predicting game outcomes from data consisting of complete games leads to overfitting. The problem is that successive positions are strongly correlated, differing by just one stone, but the regression target is shared for the entire game. When trained on the KGS data set in this way, the value network memorized the game outcomes rather than generalizing to new positions, achieving a minimum MSE of 0.37 on the test set, compared to 0.19 on the training set. To mitigate this problem, we generated a new self-play data set consisting of 30 million distinct positions, each sampled from a separate game. Each game was played between the RL policy network and itself until the game terminated. Training on this data set led to MSEs of 0.226 and 0.234 on the training and test set respectively, indicating minimal overfitting. Figure 2b shows the position evaluation accuracy of the value network, compared to Monte Carlo rollouts using the fast rollout policy p_π ; the value function was consistently more accurate. A single evaluation of $v_\theta(s)$ also approached the accuracy of Monte Carlo rollouts using the RL policy network p_σ , but using 15,000 times less computation.

Searching with policy and value networks

AlphaGo combines the policy and value networks in an MCTS algorithm (Fig. 3) that selects actions by lookahead search. Each edge

is evaluated in two ways: using the value network v_θ ; and by running a rollout to the end of the game with the fast rollout policy p_π , then computing the winner with function r . **d**, Action values Q are updated to track the mean value of all evaluations $r(\cdot)$ and $v_\theta(\cdot)$ in the subtree below that action.

(s, a) of the search tree stores an action value $Q(s, a)$, visit count $N(s, a)$, and prior probability $P(s, a)$. The tree is traversed by simulation (that is, descending the tree in complete games without backup), starting from the root state. At each time step t of each simulation, an action a_t is selected from state s_t

$$a_t = \operatorname{argmax}_a (Q(s_t, a) + u(s_t, a))$$

so as to maximize action value plus a bonus

$$u(s, a) \propto \frac{P(s, a)}{1 + N(s, a)}$$

that is proportional to the prior probability but decays with repeated visits to encourage exploration. When the traversal reaches a leaf node s_L at step L , the leaf node may be expanded. The leaf position s_L is processed just once by the SL policy network p_σ . The output probabilities are stored as prior probabilities P for each legal action a , $P(s, a) = p_\sigma(a|s)$. The leaf node is evaluated in two very different ways: first, by the value network $v_\theta(s_L)$; and second, by the outcome z_L of a random rollout played out until terminal step T using the fast rollout policy p_π ; these evaluations are combined, using a mixing parameter λ , into a leaf evaluation $V(s_L)$

$$V(s_L) = (1 - \lambda)v_\theta(s_L) + \lambda z_L$$

At the end of simulation, the action values and visit counts of all traversed edges are updated. Each edge accumulates the visit count and mean evaluation of all simulations passing through that edge

$$N(s, a) = \sum_{i=1}^n 1(s, a, i)$$

$$Q(s, a) = \frac{1}{N(s, a)} \sum_{i=1}^n 1(s, a, i) V(s_L^i)$$

where s_L^i is the leaf node from the i th simulation, and $1(s, a, i)$ indicates whether an edge (s, a) was traversed during the i th simulation. Once the search is complete, the algorithm chooses the most visited move from the root position.

It is worth noting that the SL policy network p_σ performed better in AlphaGo than the stronger RL policy network p_ρ , presumably because humans select a diverse beam of promising moves, whereas RL optimizes for the single best move. However, the value function $v_\theta(s) \approx v_{p_\sigma}(s)$ derived from the stronger RL policy network performed

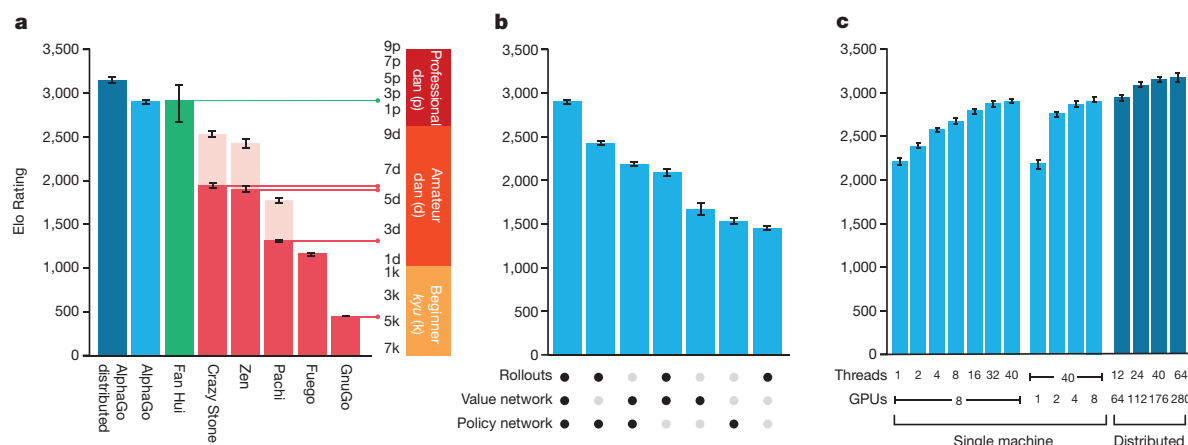


Figure 4 | Tournament evaluation of AlphaGo. **a**, Results of a tournament between different Go programs (see Extended Data Tables 6–11). Each program used approximately 5 s computation time per move. To provide a greater challenge to AlphaGo, some programs (pale upper bars) were given four handicap stones (that is, free moves at the start of every game) against all opponents. Programs were evaluated on an Elo scale³⁷: a 230 point gap corresponds to a 79% probability of winning, which roughly corresponds to one amateur *dan* rank advantage on KGS³⁸; an approximate correspondence to human ranks is also shown,

horizontal lines show KGS ranks achieved online by that program. Games against the human European champion Fan Hui were also included; these games used longer time controls. 95% confidence intervals are shown. **b**, Performance of AlphaGo, on a single machine, for different combinations of components. The version solely using the policy network does not perform any search. **c**, Scalability study of MCTS in AlphaGo with search threads and GPUs, using asynchronous search (light blue) or distributed search (dark blue), for 2 s per move.

better in AlphaGo than a value function $v_\theta(s) \approx v^{p_\theta}(s)$ derived from the SL policy network.

Evaluating policy and value networks requires several orders of magnitude more computation than traditional search heuristics. To efficiently combine MCTS with deep neural networks, AlphaGo uses an asynchronous multi-threaded search that executes simulations on CPUs, and computes policy and value networks in parallel on GPUs. The final version of AlphaGo used 40 search threads, 48 CPUs, and 8 GPUs. We also implemented a distributed version of AlphaGo that

exploited multiple machines, 40 search threads, 1,202 CPUs and 176 GPUs. The Methods section provides full details of asynchronous and distributed MCTS.

Evaluating the playing strength of AlphaGo

To evaluate AlphaGo, we ran an internal tournament among variants of AlphaGo and several other Go programs, including the strongest commercial programs Crazy Stone¹³ and Zen, and the strongest open source programs Pachi¹⁴ and Fuego¹⁵. All of these programs are based

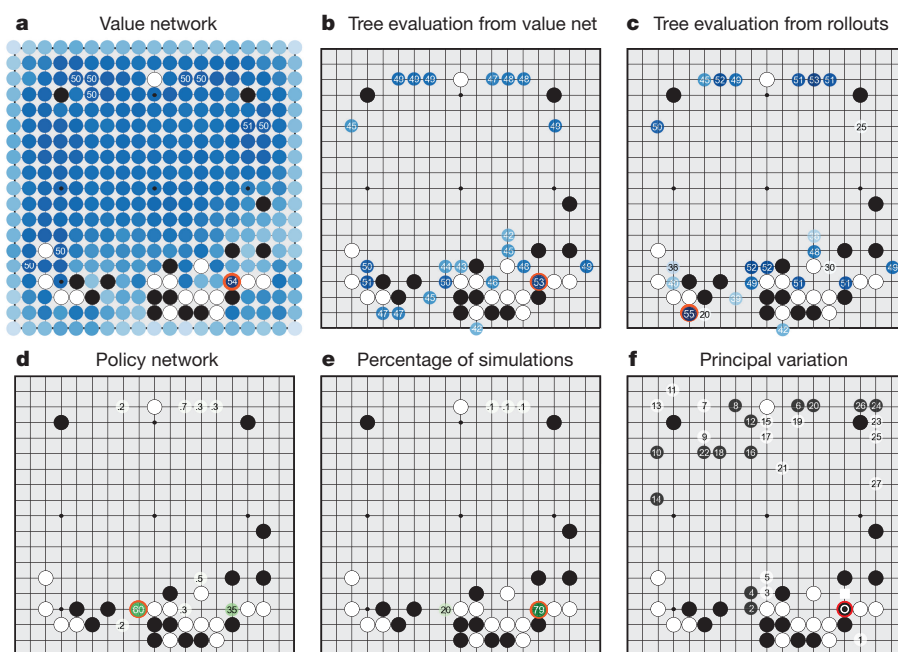


Figure 5 | How AlphaGo (black, to play) selected its move in an informal game against Fan Hui. For each of the following statistics, the location of the maximum value is indicated by an orange circle. **a**, Evaluation of all successors s' of the root position s , using the value network $v_\theta(s')$; estimated winning percentages are shown for the top evaluations. **b**, Action values $Q(s, a)$ for each edge (s, a) in the tree from root position s ; averaged over value network evaluations only ($\lambda = 0$). **c**, Action values $Q(s, a)$, averaged over rollout evaluations only ($\lambda = 1$).

d, Move probabilities directly from the SL policy network, $p_\theta(a|s)$; reported as a percentage (if above 0.1%). **e**, Percentage frequency with which actions were selected from the root during simulations. **f**, The principal variation (path with maximum visit count) from AlphaGo's search tree. The moves are presented in a numbered sequence. AlphaGo selected the move indicated by the red circle; Fan Hui responded with the move indicated by the white square; in his post-game commentary he preferred the move (labelled 1) predicted by AlphaGo.

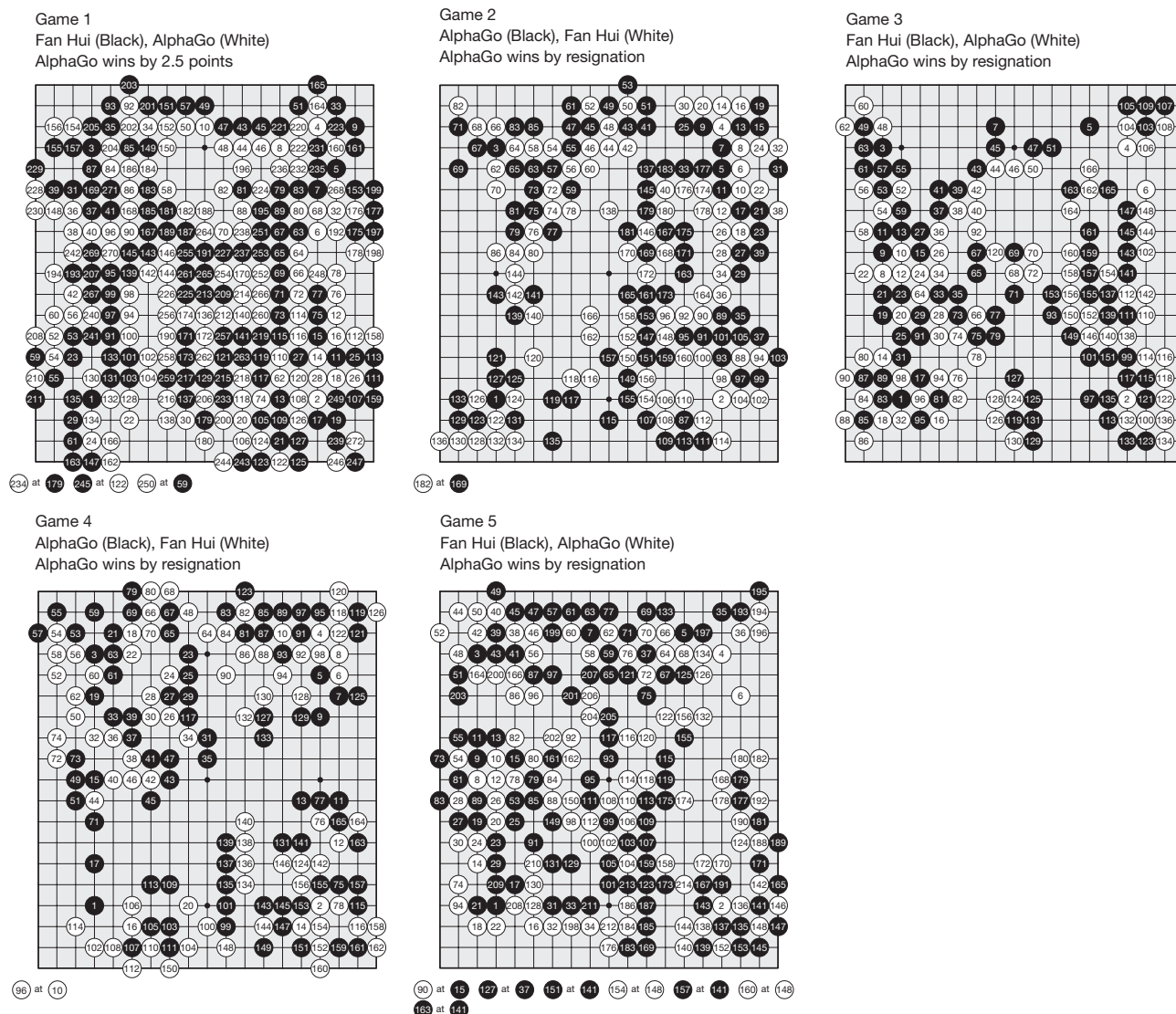


Figure 6 | Games from the match between AlphaGo and the European champion, Fan Hui. Moves are shown in a numbered sequence corresponding to the order in which they were played. Repeated moves on the same intersection are shown in pairs below the board. The first

move number in each pair indicates when the repeat move was played, at an intersection identified by the second move number (see Supplementary Information).

on high-performance MCTS algorithms. In addition, we included the open source program GnuGo, a Go program using state-of-the-art search methods that preceded MCTS. All programs were allowed 5 s of computation time per move.

The results of the tournament (see Fig. 4a) suggest that single-machine AlphaGo is many *dan* ranks stronger than any previous Go program, winning 494 out of 495 games (99.8%) against other Go programs. To provide a greater challenge to AlphaGo, we also played games with four handicap stones (that is, free moves for the opponent); AlphaGo won 77%, 86%, and 99% of handicap games against Crazy Stone, Zen and Pachi, respectively. The distributed version of AlphaGo was significantly stronger, winning 77% of games against single-machine AlphaGo and 100% of its games against other programs.

We also assessed variants of AlphaGo that evaluated positions using just the value network ($\lambda = 0$) or just rollouts ($\lambda = 1$) (see Fig. 4b). Even without rollouts AlphaGo exceeded the performance of all other Go programs, demonstrating that value networks provide a viable alternative to Monte Carlo evaluation in Go. However, the mixed evaluation ($\lambda = 0.5$) performed best, winning $\geq 95\%$ of games against other variants. This suggests that the two position-evaluation

mechanisms are complementary: the value network approximates the outcome of games played by the strong but impractically slow p_ρ , while the rollouts can precisely score and evaluate the outcome of games played by the weaker but faster rollout policy p_π . Figure 5 visualizes the evaluation of a real game position by AlphaGo.

Finally, we evaluated the distributed version of AlphaGo against Fan Hui, a professional 2 *dan*, and the winner of the 2013, 2014 and 2015 European Go championships. Over 5–9 October 2015 AlphaGo and Fan Hui competed in a formal five-game match. AlphaGo won the match 5 games to 0 (Fig. 6 and Extended Data Table 1). This is the first time that a computer Go program has defeated a human professional player, without handicap, in the full game of Go—a feat that was previously believed to be at least a decade away^{3,7,31}.

Discussion

In this work we have developed a Go program, based on a combination of deep neural networks and tree search, that plays at the level of the strongest human players, thereby achieving one of artificial intelligence's “grand challenges”^{31–33}. We have developed, for the first time, effective move selection and position evaluation functions for Go, based on deep neural networks that are trained by a novel combination

of supervised and reinforcement learning. We have introduced a new search algorithm that successfully combines neural network evaluations with Monte Carlo rollouts. Our program AlphaGo integrates these components together, at scale, in a high-performance tree search engine.

During the match against Fan Hui, AlphaGo evaluated thousands of times fewer positions than Deep Blue did in its chess match against Kasparov⁴; compensating by selecting those positions more intelligently, using the policy network, and evaluating them more precisely, using the value network—an approach that is perhaps closer to how humans play. Furthermore, while Deep Blue relied on a handcrafted evaluation function, the neural networks of AlphaGo are trained directly from gameplay purely through general-purpose supervised and reinforcement learning methods.

Go is exemplary in many ways of the difficulties faced by artificial intelligence^{33,34}: a challenging decision-making task, an intractable search space, and an optimal solution so complex it appears infeasible to directly approximate using a policy or value function. The previous major breakthrough in computer Go, the introduction of MCTS, led to corresponding advances in many other domains; for example, general game-playing, classical planning, partially observed planning, scheduling, and constraint satisfaction^{35,36}. By combining tree search with policy and value networks, AlphaGo has finally reached a professional level in Go, providing hope that human-level performance can now be achieved in other seemingly intractable artificial intelligence domains.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 November 2015; accepted 5 January 2016.

- Allis, L. V. *Searching for Solutions in Games and Artificial Intelligence*. PhD thesis, Univ. Limburg, Maastricht, The Netherlands (1994).
- van den Herik, H., Uiterwijk, J. W. & van Rijswijk, J. Games solved: now and in the future. *Artif. Intell.* **134**, 277–311 (2002).
- Schaeffer, J. The games computers (and people) play. *Advances in Computers* **52**, 189–266 (2000).
- Campbell, M., Hoane, A. & Hsu, F. Deep Blue. *Artif. Intell.* **134**, 57–83 (2002).
- Schaeffer, J. et al. A world championship caliber checkers program. *Artif. Intell.* **53**, 273–289 (1992).
- Buro, M. From simple features to sophisticated evaluation functions. In *1st International Conference on Computers and Games*, 126–145 (1999).
- Müller, M. Computer Go. *Artif. Intell.* **134**, 145–179 (2002).
- Tesauro, G. & Galperin, G. On-line policy improvement using Monte-Carlo search. In *Advances in Neural Information Processing*, 1068–1074 (1996).
- Sheppard, B. World-championship-caliber Scrabble. *Artif. Intell.* **134**, 241–275 (2002).
- Bouzy, B. & Helmstetter, B. Monte-Carlo Go developments. In *10th International Conference on Advances in Computer Games*, 159–174 (2003).
- Coulom, R. Efficient selectivity and backup operators in Monte-Carlo tree search. In *5th International Conference on Computers and Games*, 72–83 (2006).
- Kocsis, L. & Szepesvári, C. Bandit based Monte-Carlo planning. In *15th European Conference on Machine Learning*, 282–293 (2006).
- Coulom, R. Computing Elo ratings of move patterns in the game of Go. *ICGA J.* **30**, 198–208 (2007).
- Baudiš, P. & Gailly, J.-L. Pachi: State of the art open source Go program. In *Advances in Computer Games*, 24–38 (Springer, 2012).
- Müller, M., Enzenberger, M., Arneson, B. & Segal, R. Fuego – an open-source framework for board games and Go engine based on Monte-Carlo tree search. *IEEE Trans. Comput. Intell. AI in Games* **2**, 259–270 (2010).
- Gelly, S. & Silver, D. Combining online and offline learning in UCT. In *17th International Conference on Machine Learning*, 273–280 (2007).
- Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097–1105 (2012).
- Lawrence, S., Giles, C. L., Tsoi, A. C. & Back, A. D. Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.* **8**, 98–113 (1997).
- Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Stern, D., Herbrich, R. & Graepel, T. Bayesian pattern ranking for move prediction in the game of Go. In *International Conference of Machine Learning*, 873–880 (2006).
- Sutskever, I. & Nair, V. Mimicking Go experts with convolutional neural networks. In *International Conference on Artificial Neural Networks*, 101–110 (2008).
- Maddison, C. J., Huang, A., Sutskever, I. & Silver, D. Move evaluation in Go using deep convolutional neural networks. *3rd International Conference on Learning Representations* (2015).
- Clark, C. & Storkey, A. J. Training deep convolutional neural networks to play go. In *32nd International Conference on Machine Learning*, 1766–1774 (2015).
- Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.* **8**, 229–256 (1992).
- Sutton, R., McAllester, D., Singh, S. & Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems*, 1057–1063 (2000).
- Sutton, R. & Barto, A. *Reinforcement Learning: an Introduction* (MIT Press, 1998).
- Schraudolph, N. N., Dayan, P. & Sejnowski, T. J. Temporal difference learning of position evaluation in the game of Go. *Adv. Neural Inf. Process. Syst.* **6**, 817–824 (1994).
- Enzenberger, M. Evaluation in Go by a neural network using soft segmentation. In *10th Advances in Computer Games Conference*, 97–108 (2003).
- Silver, D., Sutton, R. & Müller, M. Temporal-difference search in computer Go. *Mach. Learn.* **87**, 183–219 (2012).
- Levinovitz, A. The mystery of Go, the ancient game that computers still can't win. *Wired Magazine* (2014).
- Mechner, D. All Systems Go. *The Sciences* **38**, 32–37 (1998).
- Mandziuk, J. Computational intelligence in mind games. In *Challenges for Computational Intelligence*, 407–442 (2007).
- Berliner, H. A chronology of computer chess and its literature. *Artif. Intell.* **10**, 201–214 (1978).
- Browne, C. et al. A survey of Monte-Carlo tree search methods. *IEEE Trans. Comput. Intell. AI in Games* **4**, 1–43 (2012).
- Gelly, S. et al. The grand challenge of computer Go: Monte Carlo tree search and extensions. *Commun. ACM* **55**, 106–113 (2012).
- Coulom, R. Whole-history rating: A Bayesian rating system for players of time-varying strength. In *International Conference on Computers and Games*, 113–124 (2008).
- KGS. Rating system math. <http://www.gokgs.com/help/rmath.html>.

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Fan Hui for agreeing to play against AlphaGo; T. Manning for refereeing the match; R. Munos and T. Schaul for helpful discussions and advice; A. Cain and M. Cant for work on the visuals; P. Dayan, G. Wayne, D. Kumaran, D. Purves, H. van Hasselt, A. Barreto and G. Ostrovski for reviewing the paper; and the rest of the DeepMind team for their support, ideas and encouragement.

Author Contributions A.H., G.v.d.D., J.S., I.A., M.La., A.G., T.G. and D.S. designed and implemented the search in AlphaGo. C.J.M., A.G., L.S., A.H., I.A., V.P., S.D., D.G., N.K., I.S., K.K. and D.S. designed and trained the neural networks in AlphaGo. J.S., J.N., A.H. and D.S. designed and implemented the evaluation framework for AlphaGo. D.S., M.Le., T.L., T.G., K.K. and D.H. managed and advised on the project. D.S., T.G., A.G. and D.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.S. (davidsilver@google.com) or D.H. (demishassabis@google.com).

METHODS

Problem setting. Many games of perfect information, such as chess, checkers, othello, backgammon and Go, may be defined as alternating Markov games³⁹. In these games, there is a state space \mathcal{S} (where state includes an indication of the current player to play); an action space $\mathcal{A}(s)$ defining the legal actions in any given state $s \in \mathcal{S}$; a state transition function $f(s, a, \xi)$ defining the successor state after selecting action a in state s and random input ξ (for example, dice); and finally a reward function $r^i(s)$ describing the reward received by player i in state s . We restrict our attention to two-player zero-sum games, $r^1(s) = -r^2(s) = r(s)$, with deterministic state transitions, $f(s, a, \xi) = f(s, a)$, and zero rewards except at a terminal time step T . The outcome of the game $z_t = \pm r(s_T)$ is the terminal reward at the end of the game from the perspective of the current player at time step t . A policy $p(a|s)$ is a probability distribution over legal actions $a \in \mathcal{A}(s)$. A value function is the expected outcome if all actions for both players are selected according to policy p , that is, $v^p(s) = \mathbb{E}[z_t | s_t = s, a_{t \dots T} \sim p]$. Zero-sum games have a unique optimal value function $v^*(s)$ that determines the outcome from state s following perfect play by both players,

$$v^*(s) = \begin{cases} z_T & \text{if } s = s_T, \\ \max_a - v^*(f(s, a)) & \text{otherwise} \end{cases}$$

Prior work. The optimal value function can be computed recursively by minimax (or equivalently negamax) search⁴⁰. Most games are too large for exhaustive minimax tree search; instead, the game is truncated by using an approximate value function $v(s) \approx v^*(s)$ in place of terminal rewards. Depth-first minimax search with alpha-beta pruning⁴⁰ has achieved superhuman performance in chess⁴, checkers⁵ and othello⁶, but it has not been effective in Go⁷.

Reinforcement learning can learn to approximate the optimal value function directly from games of self-play³⁹. The majority of prior work has focused on a linear combination $v_\theta(s) = \varphi(s) \cdot \theta$ of features $\varphi(s)$ with weights θ . Weights were trained using temporal-difference learning⁴¹ in chess^{42,43}, checkers^{44,45} and Go³⁰; or using linear regression in othello⁶ and Scrabble⁹. Temporal-difference learning has also been used to train a neural network to approximate the optimal value function, achieving superhuman performance in backgammon⁴⁶; and achieving weak *kyu*-level performance in small-board Go^{28,29,47} using convolutional networks.

An alternative approach to minimax search is Monte Carlo tree search (MCTS)^{11,12}, which estimates the optimal value of interior nodes by a double approximation, $V^n(s) \approx v^n(s) \approx v^*(s)$. The first approximation, $V^n(s) \approx v^n(s)$, uses n Monte Carlo simulations to estimate the value function of a simulation policy P^n . The second approximation, $v^n(s) \approx v^*(s)$, uses a simulation policy P^n in place of minimax optimal actions. The simulation policy selects actions according to a search control function $\arg\max_a (Q^n(s, a) + u(s, a))$, such as UCT¹², that selects children with higher action values, $Q^n(s, a) = -V^n(f(s, a))$, plus a bonus $u(s, a)$ that encourages exploration; or in the absence of a search tree at state s , it samples actions from a fast rollout policy $p_n(a|s)$. As more simulations are executed and the search tree grows deeper, the simulation policy becomes informed by increasingly accurate statistics. In the limit, both approximations become exact and MCTS (for example, with UCT) converges¹² to the optimal value function $\lim_{n \rightarrow \infty} V^n(s) = \lim_{n \rightarrow \infty} v^n(s) = v^*(s)$. The strongest current Go programs are based on MCTS^{13–15,36}.

MCTS has previously been combined with a policy that is used to narrow the beam of the search tree to high-probability moves¹³; or to bias the bonus term towards high-probability moves⁴⁸. MCTS has also been combined with a value function that is used to initialize action values in newly expanded nodes¹⁶, or to mix Monte Carlo evaluation with minimax evaluation⁴⁹. By contrast, AlphaGo's use of value functions is based on truncated Monte Carlo search algorithms^{8,9}, which terminate rollouts before the end of the game and use a value function in place of the terminal reward. AlphaGo's position evaluation mixes full rollouts with truncated rollouts, resembling in some respects the well-known temporal-difference learning algorithm TD(λ). AlphaGo also differs from prior work by using slower but more powerful representations of the policy and value function; evaluating deep neural networks is several orders of magnitude slower than linear representations and must therefore occur asynchronously.

The performance of MCTS is to a large degree determined by the quality of the rollout policy. Prior work has focused on handcrafted patterns⁵⁰ or learning rollout policies by supervised learning¹³, reinforcement learning¹⁶, simulation balancing^{51,52} or online adaptation^{30,53}; however, it is known that rollout-based position evaluation is frequently inaccurate⁵⁴. AlphaGo uses relatively simple rollouts, and instead addresses the challenging problem of position evaluation more directly using value networks.

Search algorithm. To efficiently integrate large neural networks into AlphaGo, we implemented an asynchronous policy and value MCTS algorithm (APV-MCTS). Each node s in the search tree contains edges (s, a) for all legal actions $a \in \mathcal{A}(s)$. Each edge stores a set of statistics,

$$\{P(s, a), N_v(s, a), N_r(s, a), W_v(s, a), W_r(s, a), Q(s, a)\}$$

where $P(s, a)$ is the prior probability, $W_v(s, a)$ and $W_r(s, a)$ are Monte Carlo estimates of total action value, accumulated over $N_v(s, a)$ and $N_r(s, a)$ leaf evaluations and rollout rewards, respectively, and $Q(s, a)$ is the combined mean action value for that edge. Multiple simulations are executed in parallel on separate search threads. The APV-MCTS algorithm proceeds in the four stages outlined in Fig. 3.

Selection (Fig. 3a). The first in-tree phase of each simulation begins at the root of the search tree and finishes when the simulation reaches a leaf node at time step L . At each of these time steps, $t < L$, an action is selected according to the statistics in the search tree, $a_t = \arg\max_a (Q(s_t, a) + u(s_t, a))$ using a variant of the PUCT

algorithm⁴⁸, $u(s, a) = c_{\text{puct}} P(s, a) \frac{\sqrt{\sum_b N_r(s, b)}}{1 + N_r(s, a)}$, where c_{puct} is a constant determining the level of exploration; this search control strategy initially prefers actions with high prior probability and low visit count, but asymptotically prefers actions with high action value.

Evaluation (Fig. 3c). The leaf position s_L is added to a queue for evaluation $v_\theta(s_L)$ by the value network, unless it has previously been evaluated. The second rollout phase of each simulation begins at leaf node s_L and continues until the end of the game. At each of these time-steps, $t \geq L$, actions are selected by both players according to the rollout policy, $a_t \sim p_\pi(\cdot | s_t)$. When the game reaches a terminal state, the outcome $z_t = \pm r(s_T)$ is computed from the final score.

Backup (Fig. 3d). At each in-tree step $t \leq L$ of the simulation, the rollout statistics are updated as if it has lost n_{vl} games, $N_r(s_t, a_t) \leftarrow N_r(s_t, a_t) + n_{\text{vl}}$; $W_r(s_t, a_t) \leftarrow W_r(s_t, a_t) - n_{\text{vl}}$; this virtual loss⁵⁵ discourages other threads from simultaneously exploring the identical variation. At the end of the simulation, the rollout statistics are updated in a backward pass through each step $t \leq L$, replacing the virtual losses by the outcome, $N_r(s_t, a_t) \leftarrow N_r(s_t, a_t) - n_{\text{vl}} + 1$; $W_r(s_t, a_t) \leftarrow W_r(s_t, a_t) + n_{\text{vl}} + z_t$. Asynchronously, a separate backward pass is initiated when the evaluation of the leaf position s_L completes. The output of the value network $v_\theta(s_L)$ is used to update value statistics in a second backward pass through each step $t \leq L$, $N_v(s_t, a_t) \leftarrow N_v(s_t, a_t) + 1$, $W_v(s_t, a_t) \leftarrow W_v(s_t, a_t) + v_\theta(s_L)$. The overall evaluation of each state action is a weighted average of the Monte Carlo estimates, $Q(s, a) = (1 - \lambda) \frac{W_v(s, a)}{N_v(s, a)} + \lambda \frac{W_r(s, a)}{N_r(s, a)}$, that mixes together the value network and rollout evaluations with weighting parameter λ . All updates are performed lock-free⁵⁶.

Expansion (Fig. 3b). When the visit count exceeds a threshold, $N_r(s, a) > n_{\text{thr}}$, the successor state $s' = f(s, a)$ is added to the search tree. The new node is initialized to $\{N(s', a) = N_r(s', a) = 0, W(s', a) = W_r(s', a) = 0, P(s', a) = p_\pi(a|s')\}$, using a tree policy $p_\pi(a|s')$ (similar to the rollout policy but with more features, see Extended Data Table 4) to provide placeholder prior probabilities for action selection. The position s' is also inserted into a queue for asynchronous GPU evaluation by the policy network. Prior probabilities are computed by the SL policy network $p_\pi^\beta(\cdot | s')$ with a softmax temperature set to β ; these replace the placeholder prior probabilities, $P(s', a) \leftarrow p_\pi^\beta(a|s')$, using an atomic update. The threshold n_{thr} is adjusted dynamically to ensure that the rate at which positions are added to the policy queue matches the rate at which the GPUs evaluate the policy network. Positions are evaluated by both the policy network and the value network using a mini-batch size of 1 to minimize end-to-end evaluation time.

We also implemented a distributed APV-MCTS algorithm. This architecture consists of a single master machine that executes the main search, many remote worker CPUs that execute asynchronous rollouts, and many remote worker GPUs that execute asynchronous policy and value network evaluations. The entire search tree is stored on the master, which only executes the in-tree phase of each simulation. The leaf positions are communicated to the worker CPUs, which execute the rollout phase of simulation, and to the worker GPUs, which compute network features and evaluate the policy and value networks. The prior probabilities of the policy network are returned to the master, where they replace placeholder prior probabilities at the newly expanded node. The rewards from rollouts and the value network outputs are each returned to the master, and backed up the originating search path.

At the end of search AlphaGo selects the action with maximum visit count; this is less sensitive to outliers than maximizing action value¹⁵. The search tree is reused at subsequent time steps: the child node corresponding to the played action becomes the new root node; the subtree below this child is retained along with all its statistics, while the remainder of the tree is discarded. The match version of AlphaGo continues searching during the opponent's move. It extends the search

if the action maximizing visit count and the action maximizing action value disagree. Time controls were otherwise shaped to use most time in the middle-game⁵⁷. AlphaGo resigns when its overall evaluation drops below an estimated 10% probability of winning the game, that is, $\max_a Q(s, a) < -0.8$.

AlphaGo does not employ the all-moves-as-first¹⁰ or rapid action value estimation⁵⁸ heuristics used in the majority of Monte Carlo Go programs; when using policy networks as prior knowledge, these biased heuristics do not appear to give any additional benefit. In addition AlphaGo does not use progressive widening¹³, dynamic *komi*⁵⁹ or an opening book⁶⁰. The parameters used by AlphaGo in the Fan Hui match are listed in Extended Data Table 5.

Rollout policy. The rollout policy $p_\pi(a|s)$ is a linear softmax policy based on fast, incrementally computed, local pattern-based features consisting of both 'response' patterns around the previous move that led to state s , and 'non-response' patterns around the candidate move a in state s . Each non-response pattern is a binary feature matching a specific 3×3 pattern centred on a , defined by the colour (black, white, empty) and liberty count (1, 2, ≥ 3) for each adjacent intersection. Each response pattern is a binary feature matching the colour and liberty count in a 12-point diamond-shaped pattern²¹ centred around the previous move. Additionally, a small number of handcrafted local features encode common-sense Go rules (see Extended Data Table 4). Similar to the policy network, the weights π of the rollout policy are trained from 8 million positions from human games on the Tygem server to maximize log likelihood by stochastic gradient descent. Rollouts execute at approximately 1,000 simulations per second per CPU thread on an empty board.

Our rollout policy $p_\pi(a|s)$ contains less handcrafted knowledge than state-of-the-art Go programs¹³. Instead, we exploit the higher-quality action selection within MCTS, which is informed both by the search tree and the policy network. We introduce a new technique that caches all moves from the search tree and then plays similar moves during rollouts; a generalization of the 'last good reply' heuristic⁵³. At every step of the tree traversal, the most probable action is inserted into a hash table, along with the 3×3 pattern context (colour, liberty and stone counts) around both the previous move and the current move. At each step of the rollout, the pattern context is matched against the hash table; if a match is found then the stored move is played with high probability.

Symmetries. In previous work, the symmetries of Go have been exploited by using rotationally and reflectionally invariant filters in the convolutional layers^{24,28,29}. Although this may be effective in small neural networks, it actually hurts performance in larger networks, as it prevents the intermediate filters from identifying specific asymmetric patterns²³. Instead, we exploit symmetries at run-time by dynamically transforming each position s using the dihedral group of eight reflections and rotations, $d_1(s), \dots, d_8(s)$. In an explicit symmetry ensemble, a mini-batch of all 8 positions is passed into the policy network or value network and computed in parallel. For the value network, the output values are simply averaged, $\bar{v}_\theta(s) = \frac{1}{8} \sum_{j=1}^8 v_\theta(d_j(s))$. For the policy network, the planes of output probabilities are rotated/reflected back into the original orientation, and averaged together to provide an ensemble prediction, $\bar{p}_\theta(\cdot|s) = \frac{1}{8} \sum_{j=1}^8 d_j^{-1}(p_\theta(\cdot|d_j(s)))$; this approach was used in our raw network evaluation (see Extended Data Table 3). Instead, APV-MCTS makes use of an implicit symmetry ensemble that randomly selects a single rotation/reflection $j \in [1, 8]$ for each evaluation. We compute exactly one evaluation for that orientation only; in each simulation we compute the value of leaf node s_L by $v_\theta(d_j(s_L))$, and allow the search procedure to average over these evaluations. Similarly, we compute the policy network for a single, randomly selected rotation/reflection, $d_j^{-1}(p_\theta(\cdot|d_j(s)))$.

Policy network: classification. We trained the policy network p_σ to classify positions according to expert moves played in the KGS data set. This data set contains 29.4 million positions from 160,000 games played by KGS 6 to 9 *dan* human players; 35.4% of the games are handicap games. The data set was split into a test set (the first million positions) and a training set (the remaining 28.4 million positions). Pass moves were excluded from the data set. Each position consisted of a raw board description s and the move a selected by the human. We augmented the data set to include all eight reflections and rotations of each position. Symmetry augmentation and input features were pre-computed for each position. For each training step, we sampled a randomly selected mini-batch of m samples from the augmented KGS data set, $\{s^k, a^k\}_{k=1}^m$ and applied an asynchronous stochastic gradient descent update to maximize the log likelihood of the action, $\Delta\sigma = \frac{\alpha}{m} \sum_{k=1}^m \frac{\partial \log p_\sigma(a^k|s^k)}{\partial \sigma}$. The step size α was initialized to 0.003 and was halved every 80 million training steps, without momentum terms, and a mini-batch size of $m = 16$. Updates were applied asynchronously on 50 GPUs using DistBelief⁶¹; gradients older than 100 steps were discarded. Training took around 3 weeks for 340 million training steps.

Policy network: reinforcement learning. We further trained the policy network by policy gradient reinforcement learning^{25,26}. Each iteration consisted of a mini-batch of n games played in parallel, between the current policy network p_ρ that is being trained, and an opponent p_ρ^- that uses parameters ρ^- from a previous iteration, randomly sampled from a pool of opponents, so as to increase the stability of training. Weights were initialized to $\rho = \rho^- = \sigma$. Every 500 iterations, we added the current parameters ρ to the opponent pool. Each game i in the mini-batch was played out until termination at step T , and then scored to determine the outcome $z_i^t = \pm r(s_{T^i})$ from each player's perspective. The games were then replayed to determine the policy gradient update, $\Delta\rho = \frac{\alpha}{n} \sum_{i=1}^n \sum_{t=1}^{T^i} \frac{\partial \log p_\rho(a_t^i|s_t^i)}{\partial \rho} (z_i^t - v(s_t^i))$, using the REINFORCE algorithm²⁵ with baseline $v(s_t^i)$ for variance reduction. On the first pass through the training pipeline, the baseline was set to zero; on the second pass we used the value network $v_\theta(s)$ as a baseline; this provided a small performance boost. The policy network was trained in this way for 10,000 mini-batches of 128 games, using 50 GPUs, for one day.

Value network: regression. We trained a value network $v_\theta(s) \approx v_{p_\rho}(s)$ to approximate the value function of the RL policy network p_ρ . To avoid overfitting to the strongly correlated positions within games, we constructed a new data set of uncorrelated self-play positions. This data set consisted of over 30 million positions, each drawn from a unique game of self-play. Each game was generated in three phases by randomly sampling a time step $U \sim \text{unif}\{1, 450\}$, and sampling the first $t = 1, \dots, U-1$ moves from the SL policy network, $a_t \sim p_\sigma(\cdot|s_t)$; then sampling one move uniformly at random from available moves, $a_U \sim \text{unif}\{1, 361\}$ (repeatedly until a_U is legal); then sampling the remaining sequence of moves until the game terminates, $t = U+1, \dots, T$, from the RL policy network, $a_t \sim p_\rho(\cdot|s_t)$. Finally, the game is scored to determine the outcome $z_t = \pm r(s_T)$. Only a single training example (s_{U+1}, z_{U+1}) is added to the data set from each game. This data provides unbiased samples of the value function $v_{p_\rho}(s_{U+1}) = \mathbb{E}[z_{U+1}|s_{U+1}, a_{U+1}, \dots, T \sim p_\rho]$. During the first two phases of generation we sample from noisier distributions so as to increase the diversity of the data set. The training method was identical to SL policy network training, except that the parameter update was based on mean squared error between the predicted values and the observed rewards,

$\Delta\theta = \frac{\alpha}{m} \sum_{k=1}^m (z^k - v_\theta(s^k)) \frac{\partial v_\theta(s^k)}{\partial \theta}$. The value network was trained for 50 million mini-batches of 32 positions, using 50 GPUs, for one week.

Features for policy/value network. Each position s was pre-processed into a set of 19×19 feature planes. The features that we use come directly from the raw representation of the game rules, indicating the status of each intersection of the Go board: stone colour, liberties (adjacent empty points of stone's chain), captures, legality, turns since stone was played, and (for the value network only) the current colour to play. In addition, we use one simple tactical feature that computes the outcome of a ladder search⁷. All features were computed relative to the current colour to play; for example, the stone colour at each intersection was represented as either player or opponent rather than black or white. Each integer feature value is split into multiple 19×19 planes of binary values (one-hot encoding). For example, separate binary feature planes are used to represent whether an intersection has 1 liberty, 2 liberties, $\dots, \geq 8$ liberties. The full set of feature planes are listed in Extended Data Table 2.

Neural network architecture. The input to the policy network is a $19 \times 19 \times 48$ image stack consisting of 48 feature planes. The first hidden layer zero pads the input into a 23×23 image, then convolves k filters of kernel size 5×5 with stride 1 with the input image and applies a rectifier nonlinearity. Each of the subsequent hidden layers 2 to 12 zero pads the respective previous hidden layer into a 21×21 image, then convolves k filters of kernel size 3×3 with stride 1, again followed by a rectifier nonlinearity. The final layer convolves 1 filter of kernel size 1×1 with stride 1, with a different bias for each position, and applies a softmax function. The match version of AlphaGo used $k = 192$ filters; Fig. 2b and Extended Data Table 3 additionally show the results of training with $k = 128, 256$ and 384 filters.

The input to the value network is also a $19 \times 19 \times 48$ image stack, with an additional binary feature plane describing the current colour to play. Hidden layers 2 to 11 are identical to the policy network, hidden layer 12 is an additional convolution layer, hidden layer 13 convolves 1 filter of kernel size 1×1 with stride 1, and hidden layer 14 is a fully connected linear layer with 256 rectifier units. The output layer is a fully connected linear layer with a single tanh unit.

Evaluation. We evaluated the relative strength of computer Go programs by running an internal tournament and measuring the Elo rating of each program. We estimate the probability that program a will beat program b by a logistic function $p(a \text{ beats } b) = \frac{1}{1 + \exp(c_{\text{elo}}(e(b) - e(a)))}$, and estimate the ratings $e(\cdot)$ by Bayesian logistic regression, computed by the *BayesElo* program³⁷ using the standard constant $c_{\text{elo}} = 1/400$. The scale was anchored to the *BayesElo* rating of professional

Go player Fan Hui (2,908 at date of submission)⁶². All programs received a maximum of 5 s computation time per move; games were scored using Chinese rules with a *komi* of 7.5 points (extra points to compensate white for playing second). We also played handicap games where AlphaGo played white against existing Go programs; for these games we used a non-standard handicap system in which *komi* was retained but black was given additional stones on the usual handicap points. Using these rules, a handicap of K stones is equivalent to giving $K - 1$ free moves to black, rather than $K - 1/2$ free moves using standard no-*komi* handicap rules. We used these handicap rules because AlphaGo's value network was trained specifically to use a *komi* of 7.5.

With the exception of distributed AlphaGo, each computer Go program was executed on its own single machine, with identical specifications, using the latest available version and the best hardware configuration supported by that program (see Extended Data Table 6). In Fig. 4, approximate ranks of computer programs are based on the highest KGS rank achieved by that program; however, the KGS version may differ from the publicly available version.

The match against Fan Hui was arbitrated by an impartial referee. Five formal games and five informal games were played with 7.5 *komi*, no handicap, and Chinese rules. AlphaGo won these games 5–0 and 3–2 respectively (Fig. 6 and Extended Data Table 1). Time controls for formal games were 1 h main time plus three periods of 30 s *byoyomi*. Time controls for informal games were three periods of 30 s *byoyomi*. Time controls and playing conditions were chosen by Fan Hui in advance of the match; it was also agreed that the overall match outcome would be determined solely by the formal games. To approximately assess the relative rating of Fan Hui to computer Go programs, we appended the results of all ten games to our internal tournament results, ignoring differences in time controls.

39. Littman, M. L. Markov games as a framework for multi-agent reinforcement learning. In *11th International Conference on Machine Learning*, 157–163 (1994).
40. Knuth, D. E. & Moore, R. W. An analysis of alpha-beta pruning. *Artif. Intell.* **6**, 293–326 (1975).
41. Sutton, R. Learning to predict by the method of temporal differences. *Mach. Learn.* **3**, 9–44 (1988).
42. Baxter, J., Tridgell, A. & Weaver, L. Learning to play chess using temporal differences. *Mach. Learn.* **40**, 243–263 (2000).
43. Veness, J., Silver, D., Blair, A. & Uther, W. Bootstrapping from game tree search. In *Advances in Neural Information Processing Systems* (2009).
44. Samuel, A. L. Some studies in machine learning using the game of checkers II - recent progress. *IBM J. Res. Develop.* **11**, 601–617 (1967).
45. Schaeffer, J., Hlynka, M. & Jussila, V. Temporal difference learning applied to a high-performance game-playing program. In *17th International Joint Conference on Artificial Intelligence*, 529–534 (2001).
46. Tesauro, G. TD-gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput.* **6**, 215–219 (1994).
47. Dahl, F. Honte, a Go-playing program using neural nets. In *Machines that learn to play games*, 205–223 (Nova Science, 1999).
48. Rosin, C. D. Multi-armed bandits with episode context. *Ann. Math. Artif. Intell.* **61**, 203–230 (2011).
49. Lanctot, M., Winands, M. H. M., Pepels, T. & Sturtevant, N. R. Monte Carlo tree search with heuristic evaluations using implicit minimax backups. In *IEEE Conference on Computational Intelligence and Games*, 1–8 (2014).
50. Gelly, S., Wang, Y., Munos, R. & Teytaud, O. Modification of UCT with patterns in Monte-Carlo Go. Tech. Rep. 6062, INRIA (2006).
51. Silver, D. & Tesauro, G. Monte-Carlo simulation balancing. In *26th International Conference on Machine Learning*, 119 (2009).
52. Huang, S.-C., Coulom, R. & Lin, S.-S. Monte-Carlo simulation balancing in practice. In *7th International Conference on Computers and Games*, 81–92 (Springer-Verlag, 2011).
53. Baier, H. & Drake, P. D. The power of forgetting: improving the last-good-reply policy in Monte Carlo Go. *IEEE Trans. Comput. Intell. AI in Games* **2**, 303–309 (2010).
54. Huang, S. & Müller, M. Investigating the limits of Monte-Carlo tree search methods in computer Go. In *8th International Conference on Computers and Games*, 39–48 (2013).
55. Segal, R. B. On the scalability of parallel UCT. *Computers and Games* **6515**, 36–47 (2011).
56. Enzenberger, M. & Müller, M. A lock-free multithreaded Monte-Carlo tree search algorithm. In *12th Advances in Computer Games Conference*, 14–20 (2009).
57. Huang, S.-C., Coulom, R. & Lin, S.-S. Time management for Monte-Carlo tree search applied to the game of Go. In *International Conference on Technologies and Applications of Artificial Intelligence*, 462–466 (2010).
58. Gelly, S. & Silver, D. Monte-Carlo tree search and rapid action value estimation in computer Go. *Artif. Intell.* **175**, 1856–1875 (2011).
59. Baudiš, P. Balancing MCTS by dynamically adjusting the komi value. *ICGA J.* **34**, 131 (2011).
60. Baier, H. & Winands, M. H. Active opening book application for Monte-Carlo tree search in 19×19 Go. In *Benelux Conference on Artificial Intelligence*, 3–10 (2011).
61. Dean, J. et al. Large scale distributed deep networks. In *Advances in Neural Information Processing Systems*, 1223–1231 (2012).
62. Go ratings. <http://www.goratings.org>.

Extended Data Table 1 | Details of match between AlphaGo and Fan Hui

Date	Black	White	Category	Result
5/10/15	Fan Hui	<i>AlphaGo</i>	Formal	<i>AlphaGo</i> wins by 2.5 points
5/10/15	Fan Hui	<i>AlphaGo</i>	Informal	Fan Hui wins by resignation
6/10/15	<i>AlphaGo</i>	Fan Hui	Formal	<i>AlphaGo</i> wins by resignation
6/10/15	<i>AlphaGo</i>	Fan Hui	Informal	<i>AlphaGo</i> wins by resignation
7/10/15	Fan Hui	<i>AlphaGo</i>	Formal	<i>AlphaGo</i> wins by resignation
7/10/15	Fan Hui	<i>AlphaGo</i>	Informal	<i>AlphaGo</i> wins by resignation
8/10/15	<i>AlphaGo</i>	Fan Hui	Formal	<i>AlphaGo</i> wins by resignation
8/10/15	<i>AlphaGo</i>	Fan Hui	Informal	<i>AlphaGo</i> wins by resignation
9/10/15	Fan Hui	<i>AlphaGo</i>	Formal	<i>AlphaGo</i> wins by resignation
9/10/15	<i>AlphaGo</i>	Fan Hui	Informal	Fan Hui wins by resignation

The match consisted of five formal games with longer time controls, and five informal games with shorter time controls. Time controls and playing conditions were chosen by Fan Hui in advance of the match.

Extended Data Table 2 | Input features for neural networks

Feature	# of planes	Description
Stone colour	3	Player stone / opponent stone / empty
Ones	1	A constant plane filled with 1
Turns since	8	How many turns since a move was played
Liberties	8	Number of liberties (empty adjacent points)
Capture size	8	How many opponent stones would be captured
Self-atari size	8	How many of own stones would be captured
Liberties after move	8	Number of liberties after this move is played
Ladder capture	1	Whether a move at this point is a successful ladder capture
Ladder escape	1	Whether a move at this point is a successful ladder escape
Sensibleness	1	Whether a move is legal and does not fill its own eyes
Zeros	1	A constant plane filled with 0
Player color	1	Whether current player is black

Feature planes used by the policy network (all but last feature) and value network (all features).

Extended Data Table 3 | Supervised learning results for the policy network

Architecture			Evaluation					
Filters	Symmetries	Features	Test accu- racy %	Train accu- racy %	Raw wins %	net wins %	<i>AlphaGo</i> wins %	Forward time (ms)
128	1	48	54.6	57.0	36		53	2.8
192	1	48	55.4	58.0	50		50	4.8
256	1	48	55.9	59.1	67		55	7.1
256	2	48	56.5	59.8	67		38	13.9
256	4	48	56.9	60.2	69		14	27.6
256	8	48	57.0	60.4	69		5	55.3
192	1	4	47.6	51.4	25		15	4.8
192	1	12	54.7	57.1	30		34	4.8
192	1	20	54.7	57.2	38		40	4.8
192	8	4	49.2	53.2	24		2	36.8
192	8	12	55.7	58.3	32		3	36.8
192	8	20	55.8	58.4	42		3	36.8

The policy network architecture consists of 128, 192 or 256 filters in convolutional layers; an explicit symmetry ensemble over 2, 4 or 8 symmetries; using only the first 4, 12 or 20 input feature planes listed in Extended Data Table 1. The results consist of the test and train accuracy on the KGS data set; and the percentage of games won by given policy network against AlphaGo's policy network (highlighted row 2): using the policy networks to select moves directly (raw wins); or using AlphaGo's search to select moves (AlphaGo wins); and finally the computation time for a single evaluation of the policy network.

Extended Data Table 4 | Input features for rollout and tree policy

Feature	# of patterns	Description
Response	1	Whether move matches one or more response pattern features
Save atari	1	Move saves stone(s) from capture
Neighbour	8	Move is 8-connected to previous move
Nakade	8192	Move matches a <i>nakade</i> pattern at captured stone
Response pattern	32207	Move matches 12-point diamond pattern near previous move
Non-response pattern	69338	Move matches 3×3 pattern around move
Self-atari	1	Move allows stones to be captured
Last move distance	34	Manhattan distance to previous two moves
Non-response pattern	32207	Move matches 12-point diamond pattern centred around move

Features used by the rollout policy (first set) and tree policy (first and second set). Patterns are based on stone colour (black/white/empty) and liberties (1, 2, ≥ 3) at each intersection of the pattern.

Extended Data Table 5 | Parameters used by AlphaGo

Symbol	Parameter	Value
β	Softmax temperature	0.67
λ	Mixing parameter	0.5
n_{vl}	Virtual loss	3
n_{thr}	Expansion threshold	40
c_{puct}	Exploration constant	5

Extended Data Table 6 | Results of a tournament between different Go programs

Short name	Computer Player	Version	Time settings	CPUs	GPUs	KGS Rank	Elo
α_{rvp}^d	Distributed AlphaGo	See Methods	5 seconds	1202	176	–	3140
α_{rvp}	AlphaGo	See Methods	5 seconds	48	8	–	2890
<i>CS</i>	CrazyStone	2015	5 seconds	32	–	6d	1929
<i>ZN</i>	Zen	5	5 seconds	8	–	6d	1888
<i>PC</i>	Pachi	10.99	400,000 sims	16	–	2d	1298
<i>FG</i>	Fuego	svn1989	100,000 sims	16	–	–	1148
<i>GG</i>	GnuGo	3.8	level 10	1	–	5k	431
CS_4	CrazyStone	4 handicap stones	5 seconds	32	–	–	2526
ZN_4	Zen	4 handicap stones	5 seconds	8	–	–	2413
PC_4	Pachi	4 handicap stones	400,000 sims	16	–	–	1756

Each program played with a maximum of 5 s thinking time per move; the games against Fan Hui were conducted using longer time controls, as described in Methods. CN_4 , ZN_4 and PC_4 were given 4 handicap stones; *komi* was 7.5 in all games. Elo ratings were computed by BayesElo.

Extended Data Table 7 | Results of a tournament between different variants of AlphaGo

Short name	Policy network	Value network	Rollouts	Mixing constant	Policy GPUs	Value GPUs	Elo rating
α_{rvp}	p_σ	v_θ	p_π	$\lambda = 0.5$	2	6	2890
α_{vp}	p_σ	v_θ	—	$\lambda = 0$	2	6	2177
α_{rp}	p_σ	—	p_π	$\lambda = 1$	8	0	2416
α_{rv}	$[p_\tau]$	v_θ	p_π	$\lambda = 0.5$	0	8	2077
α_v	$[p_\tau]$	v_θ	—	$\lambda = 0$	0	8	1655
α_r	$[p_\tau]$	—	p_π	$\lambda = 1$	0	0	1457
α_p	p_σ	—	—	—	0	0	1517

Evaluating positions using rollouts only (α_{rp} , α_r), value nets only (α_{vp} , α_v), or mixing both (α_{rvp} , α_{rv}); either using the policy network $p_\sigma(\alpha_{rvp}$, α_{vp} , α_{rp}), or no policy network (α_{rvp} , α_{vp} , α_{rp}), that is, instead using the placeholder probabilities from the tree policy p_τ throughout. Each program used 5 s per move on a single machine with 48 CPUs and 8 GPUs. Elo ratings were computed by BayesElo.

Extended Data Table 8 | Results of a tournament between AlphaGo and distributed AlphaGo, testing scalability with hardware

<i>AlphaGo</i>	Search threads	CPUs	GPUs	Elo
Asynchronous	1	48	8	2203
Asynchronous	2	48	8	2393
Asynchronous	4	48	8	2564
Asynchronous	8	48	8	2665
Asynchronous	16	48	8	2778
Asynchronous	32	48	8	2867
Asynchronous	40	48	8	2890
Asynchronous	40	48	1	2181
Asynchronous	40	48	2	2738
Asynchronous	40	48	4	2850
Distributed	12	428	64	2937
Distributed	24	764	112	3079
Distributed	40	1202	176	3140
Distributed	64	1920	280	3168

Each program played with a maximum of 2 s thinking time per move. Elo ratings were computed by BayesElo.

Extended Data Table 9 | Cross-table of win rates in per cent between programs

	α_{rvp}	α_{vp}	α_{rp}	α_{rv}	α_r	α_v	α_p
α_{rvp}	-	1 [0; 5]	5 [4; 7]	0 [0; 4]	0 [0; 8]	0 [0; 19]	0 [0; 19]
α_{vp}	99 [95; 100]	-	61 [52; 69]	35 [25; 48]	6 [1; 27]	0 [0; 22]	1 [0; 6]
α_{rp}	95 [93; 96]	39 [31; 48]	-	13 [7; 23]	0 [0; 9]	0 [0; 22]	4 [1; 21]
α_{rv}	100 [96; 100]	65 [52; 75]	87 [77; 93]	-	0 [0; 18]	29 [8; 64]	48 [33; 65]
α_r	100 [92; 100]	94 [73; 99]	100 [91; 100]	100 [82; 100]	-	78 [45; 94]	78 [71; 84]
α_v	100 [81; 100]	100 [78; 100]	100 [78; 100]	71 [36; 92]	22 [6; 55]	-	30 [16; 48]
α_p	100 [81; 100]	99 [94; 100]	96 [79; 99]	52 [35; 67]	22 [16; 29]	70 [52; 84]	-
CS	100 [97; 100]	74 [66; 81]	98 [94; 99]	80 [70; 87]	5 [3; 7]	36 [16; 61]	8 [5; 14]
ZN	99 [93; 100]	84 [67; 93]	98 [93; 99]	92 [67; 99]	6 [2; 19]	40 [12; 77]	100 [65; 100]
PC	100 [98; 100]	99 [95; 100]	100 [98; 100]	98 [89; 100]	78 [73; 81]	87 [68; 95]	55 [47; 62]
FG	100 [97; 100]	99 [93; 100]	100 [96; 100]	100 [91; 100]	78 [73; 83]	100 [65; 100]	65 [55; 73]
GG	100 [44; 100]	100 [34; 100]	100 [68; 100]	100 [57; 100]	99 [97; 100]	67 [21; 94]	99 [95; 100]
CS_4	77 [69; 84]	12 [8; 18]	53 [44; 61]	15 [8; 24]	0 [0; 3]	0 [0; 30]	0 [0; 8]
ZN_4	86 [77; 92]	25 [16; 38]	67 [56; 76]	14 [7; 27]	0 [0; 12]	0 [0; 43]	-
PC_4	99 [97; 100]	82 [75; 88]	98 [95; 99]	89 [79; 95]	32 [26; 39]	13 [3; 36]	35 [25; 46]

95% Agresti–Coull confidence intervals in grey. Each program played with a maximum of 5 s thinking time per move. CN_4 , ZN_4 and PC_4 were given 4 handicap stones; *komi* was 7.5 in all games. Distributed AlphaGo scored 77% [70; 82] against α_{vp} and 100% against all other programs (no handicap games were played).

Extended Data Table 10 | Cross-table of win rates in per cent between programs in the single-machine scalability study

Threads	GPU	1	2	4	8	16	32	40	40	40	40
		8	8	8	8	8	8	8	4	2	1
1	8	-	70 [61;78]	90 [84;94]	94 [83;98]	86 [72;94]	98 [91;100]	98 [92;99]	100 [76;100]	96 [91;98]	38 [25;52]
2	8	30 [22;39]	-	72 [61;81]	81 [71;88]	86 [76;93]	92 [83;97]	93 [86;96]	83 [69;91]	84 [75;90]	26 [17;38]
4	8	10 [6;16]	28 [19;39]	-	62 [53;70]	71 [61;80]	82 [71;89]	84 [74;90]	81 [69;89]	78 [63;88]	18 [10;28]
8	8	6 [2;17]	19 [12;29]	38 [30;47]	-	61 [51;71]	65 [51;76]	73 [62;82]	74 [59;85]	64 [55;73]	12 [3;34]
16	8	14 [6;28]	14 [7;24]	29 [20;39]	39 [29;49]	-	52 [41;63]	61 [50;71]	52 [41;64]	41 [32;51]	5 [1;25]
32	8	2 [0;9]	8 [3;17]	18 [11;29]	35 [24;49]	48 [37;59]	-	52 [42;63]	44 [32;57]	26 [17;36]	0 [0;30]
40	8	2 [1;8]	8 [4;14]	16 [10;26]	27 [18;38]	39 [29;50]	48 [37;58]	-	43 [30;56]	41 [26;58]	4 [1;18]
40	4	0 [0;24]	17 [9;31]	19 [11;31]	26 [15;41]	48 [36;59]	56 [43;68]	57 [44;70]	-	29 [18;41]	2 [0;11]
40	2	4 [2;9]	16 [10;25]	22 [12;37]	36 [27;45]	59 [49;68]	74 [64;83]	59 [42;74]	71 [59;82]	-	5 [1;17]
40	1	62 [48;75]	74 [62;83]	82 [72;90]	88 [66;97]	95 [75;99]	100 [70;100]	96 [82;99]	98 [89;100]	95 [83;99]	-

95% Agresti–Coull confidence intervals in grey. Each program played with 2 s per move; *komi* was 7.5 in all games.

Extended Data Table 11 | Cross-table of win rates in per cent between programs in the distributed scalability study

Threads			40	12	24	40	64
GPU			8	64	112	176	280
CPU			48	428	764	1202	1920
40	8	48	-	52 [43; 61]	68 [59; 76]	77 [70; 82]	81 [65; 91]
12	64	428	48 [39; 57]	-	64 [54; 73]	62 [41; 79]	83 [55; 95]
24	112	764	32 [24; 41]	36 [27; 46]	-	36 [20; 57]	60 [51; 69]
40	176	1202	23 [18; 30]	38 [21; 59]	64 [43; 80]	-	53 [39; 67]
64	280	1920	19 [9; 35]	17 [5; 45]	40 [31; 49]	47 [33; 61]	-

95% Agresti–Coull confidence intervals in grey. Each program played with 2 s per move; *komi* was 7.5 in all games.

High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects

Benjamin P. Kleinstiver^{1,2*}, Vikram Pattanayak^{1,2*}, Michelle S. Prew¹, Shengdar Q. Tsai^{1,2}, Nhu T. Nguyen¹, Zongli Zheng³ & J. Keith Joung^{1,2}

CRISPR–Cas9 nucleases are widely used for genome editing but can induce unwanted off-target mutations. Existing strategies for reducing genome-wide off-target effects of the widely used *Streptococcus pyogenes* Cas9 (SpCas9) are imperfect, possessing only partial or unproven efficacies and other limitations that constrain their use. Here we describe SpCas9–HF1, a high-fidelity variant harbouring alterations designed to reduce non-specific DNA contacts. SpCas9–HF1 retains on-target activities comparable to wild-type SpCas9 with >85% of single-guide RNAs (sgRNAs) tested in human cells. Notably, with sgRNAs targeted to standard non-repetitive sequences, SpCas9–HF1 rendered all or nearly all off-target events undetectable by genome-wide break capture and targeted sequencing methods. Even for atypical, repetitive target sites, the vast majority of off-target mutations induced by wild-type SpCas9 were not detected with SpCas9–HF1. With its exceptional precision, SpCas9–HF1 provides an alternative to wild-type SpCas9 for research and therapeutic applications. More broadly, our results suggest a general strategy for optimizing genome-wide specificities of other CRISPR–RNA-guided nucleases.

CRISPR–Cas9 nucleases enable highly efficient genome editing in a wide variety of organisms^{1–3}, but can also cause unwanted mutations at off-target sites that resemble the on-target sequence^{4–13}. These off-target effects can confound research experiments and also have potential implications for therapeutic uses of the technology. Various strategies have been described to reduce genome-wide off-target mutations of the commonly used SpCas9 nuclease, including: truncated sgRNAs bearing shortened regions of target site complementarity^{8,14}, SpCas9 mutants such as the recently described D1135E variant¹⁵, paired SpCas9 nickases^{16,17}, and dimeric fusions of catalytically inactive SpCas9 to a non-specific FokI nuclease^{18–20}. However, these approaches are only partially effective, have as-yet unproven efficacies on a genome-wide scale, and/or possess the potential to create more new off-target sites. Furthermore, some require expression of multiple sgRNAs and/or fusion of additional functional domains to Cas9, which can reduce targeting range and create challenges for delivery with viral vectors that have limits on nucleic acid payload size. Thus, a major challenge for the field remains the development of a robust and easily used strategy that eliminates off-target mutations on a genome-wide scale.

We initially hypothesized that off-target effects of SpCas9 might be minimized by decreasing non-specific interactions with its target DNA site. SpCas9–sgRNA complexes cleave target sites composed of an NGG protospacer adjacent motif (PAM) sequence (recognized by SpCas9)^{21–24} and an adjacent 20 base pair (bp) protospacer sequence (which is complementary to the 5' end of the sgRNA)^{22,25–27}. We previously proposed that the SpCas9–sgRNA complex might possess more energy than is needed for optimal recognition of its intended target DNA site, thereby enabling cleavage of mismatched off-target sites¹⁴. Structural studies have suggested that the SpCas9–sgRNA–target DNA complex encompasses several SpCas9-mediated DNA contacts, including direct hydrogen bonds made by four SpCas9

residues (N497, R661, Q695, Q926) to the phosphate backbone of the target DNA strand^{28,29} (Fig. 1a and Extended Data Fig. 1a, b). We envisioned that disruption of one or more of these contacts might alter the energetics of the SpCas9–sgRNA complex so that it might retain enough for robust on-target activity but have a diminished ability to cleave mismatched off-target sites.

Alteration of SpCas9 DNA contacts

Guided by this excess energy hypothesis, we first constructed 15 different SpCas9 variants bearing all possible single, double, triple, and quadruple combinations of N497A, R661A, Q695A, and Q926A substitutions to test whether contacts made by these residues might be dispensable for on-target activity (Fig. 1b). For these experiments, we used a previously described human cell-based enhanced GFP (EGFP) disruption assay³⁰. Using an EGFP-targeted sgRNA, which we have previously shown can efficiently induce insertion or deletion mutations (indels) in an EGFP reporter gene when paired with wild-type SpCas9 (ref. 4), we found that all 15 SpCas9 variants possessed activities comparable to that of wild-type SpCas9 (Fig. 1b, grey bars). Thus, alanine substitution of one or all of these residues did not reduce on-target cleavage efficiency of SpCas9 with this EGFP-targeted sgRNA.

Next, we sought to assess the relative activities of all 15 SpCas9 variants at mismatched target sites. To do this, we repeated the EGFP disruption assay with derivatives of the EGFP-targeted sgRNA used in the previous experiment that contain pairs of substituted bases at positions ranging from 13 to 19 (numbering starting with 1 for the most PAM-proximal base and ending with 20 for the most PAM-distal base; Fig. 1b). This analysis revealed that one of the triply substituted variants (R661A/Q695A/Q926A) and the quadruple substitution variant (N497A/R661A/Q695A/Q926A) both showed minimal EGFP disruption at or near background levels with all four of the mismatched

¹Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA.

²Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China.

*These authors contributed equally to this work.

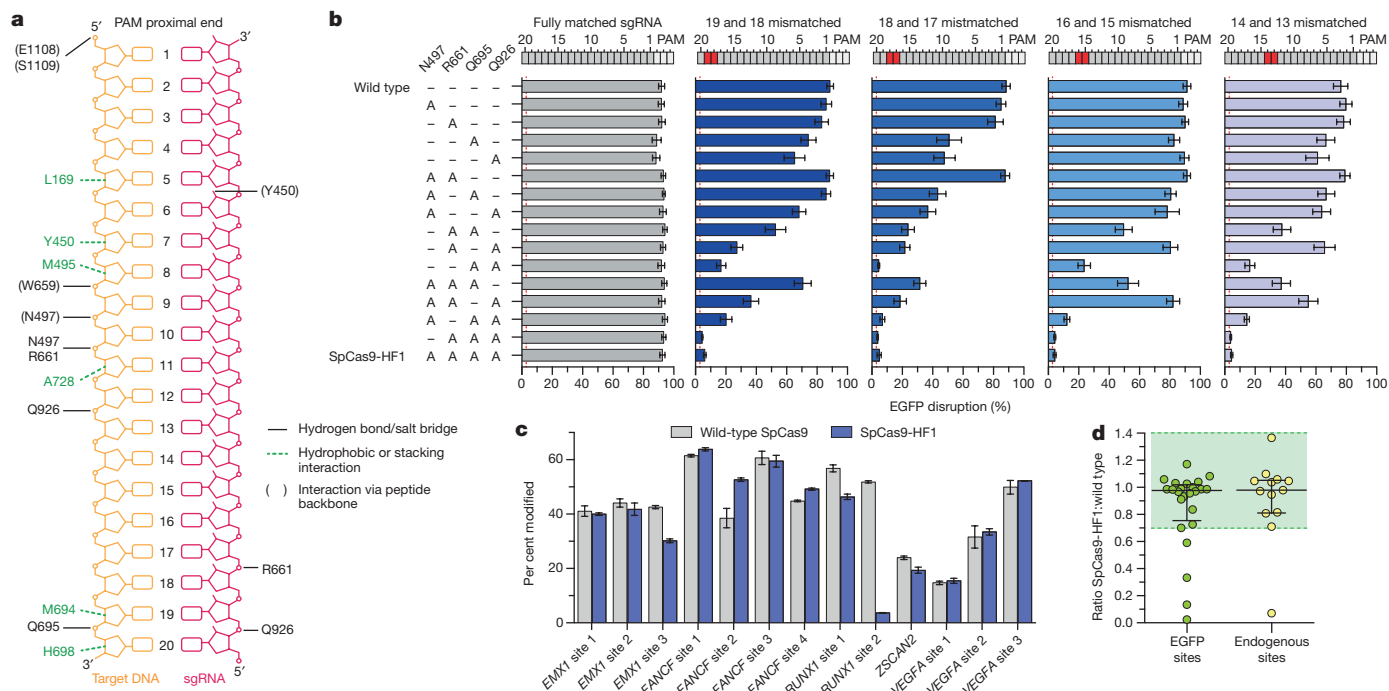


Figure 1 | Identification and characterization of SpCas9 variants bearing substitutions in residues that form non-specific DNA contacts. **a**, Schematic depicting wild-type SpCas9 interactions with the target DNA-sgRNA duplex, based on PDB accession 4O08 and 4UN3 (adapted from refs 28 and 29, respectively). **b**, Characterization of SpCas9 variants that contain alanine substitutions in positions that form hydrogen bonds with the DNA backbone. Wild-type SpCas9 and variants were assessed using the human cell EGFP disruption assay when programmed with a

sgRNAs (Fig. 1b, coloured bars). Based on these results, we chose the quadruple substitution variant (hereafter referred to as SpCas9-HF1 for high-fidelity variant number 1) for further analysis.

SpCas9-HF1 retains high on-target activities

To determine how robustly SpCas9-HF1 functions at a larger number of on-target sites, we performed direct comparisons between this variant and wild-type SpCas9 using additional sgRNAs. In total, we tested 37 different sgRNAs, 24 targeted to *EGFP* and 13 targeted to endogenous human gene targets. For 20 of the 24 sgRNAs tested using the EGFP disruption assay (Extended Data Fig. 2a) and 12 of the 13 sgRNAs tested using a T7 endonuclease I mismatch assay (Fig. 1c), we found SpCas9-HF1 exhibited at least 70% of the on-target activities observed with wild-type SpCas9 at the same sites (Fig. 1d). Indeed, SpCas9-HF1 showed highly comparable activities (90–140%) to wild-type SpCas9 with the vast majority of sgRNAs (Fig. 1d). Three of the 37 sgRNAs tested showed essentially no activity with SpCas9-HF1 (*EGFP* sites 9 and 23, and *RUNX1* site 2), and examination of these target sites did not suggest any obvious differences in the characteristics of these sequences compared to those for which we saw high activities (Supplementary Table 1). Overall, SpCas9-HF1 possesses comparable activities (greater than 70% of wild-type SpCas9 activities) for 86% (32/37) of the sgRNAs we tested.

Genome-wide specificity of SpCas9-HF1

To test whether SpCas9-HF1 exhibits reduced off-target effects in human cells, we used the genome-wide unbiased identification of double-stranded breaks enabled by sequencing (GUIDE-seq) method⁸ to assess eight different sgRNAs targeted to sites in the endogenous human *EMX1*, *FANCF*, *RUNX1*, and *ZSCAN2* genes. The sequences targeted by these sgRNAs have variable numbers of predicted mismatched sites in the reference human genome (Extended Data Table 1). Assessment

perfectly matched sgRNA or partially mismatched sgRNAs. Error bars represent s.e.m. for $n = 3$; mean level of background EGFP loss represented by red dashed line. **c**, On-target activities of wild-type SpCas9 and SpCas9-HF1 across 13 endogenous sites measured by T7 endonuclease I assay. Error bars represent s.e.m. for $n = 3$. **d**, Ratio of on-target activity of SpCas9-HF1 to wild-type SpCas9. The median and interquartile range are shown; the interval with $>70\%$ of wild-type activity is highlighted in green.

of on-target double-stranded oligodeoxynucleotide (dsODN) tag integration (by restriction-fragment length polymorphism (RFLP) assay) and indel formation (by T7 endonuclease I assay) for the eight sgRNAs revealed comparable on-target activities with wild-type SpCas9 and SpCas9-HF1 (Extended Data Fig. 3a and 3b, respectively), demonstrating that these GUIDE-seq experiments were working efficiently and comparably with the two different nucleases.

These GUIDE-seq experiments showed that with wild-type SpCas9, seven of the eight sgRNAs induced cleavage at multiple off-target sites (ranging from 2 to 25 per sgRNA), whereas the eighth sgRNA (*FANCF* site 4) did not yield any detectable off-target sites (Fig. 2a, b). The off-target sites identified harboured one to six mismatches distributed throughout various positions in the protospacer and/or PAM sequence (Fig. 2c and Extended Data Fig. 4a). However, with SpCas9-HF1, a complete absence of GUIDE-seq detectable off-target events was observed for six of the seven sgRNAs that induced off-target effects with wild-type SpCas9 (Fig. 2a, b). Among these seven sgRNAs, only a single detectable genome-wide off-target was identified, for *FANCF* site 2, at a site harbouring one mismatch within the protospacer seed sequence (Fig. 2a). As with wild-type SpCas9, the eighth sgRNA (*FANCF* site 4) did not yield any detectable off-target cleavage events when tested with SpCas9-HF1 (Fig. 2a). Notably, with all eight sgRNAs, SpCas9-HF1 did not create any new nuclease-induced off-target sites (not already observed with wild-type SpCas9) detectable by GUIDE-seq.

To confirm these GUIDE-seq findings, we used targeted amplicon sequencing to more directly measure the frequencies of indel mutations induced by wild-type SpCas9 and SpCas9-HF1. For these experiments, we transfected human cells only with sgRNA- and Cas9-encoding plasmids (without the GUIDE-seq tag). We used next-generation sequencing to examine the on-target sites and 36 of the 40 off-target sites that had been identified for six sgRNAs with wild-type

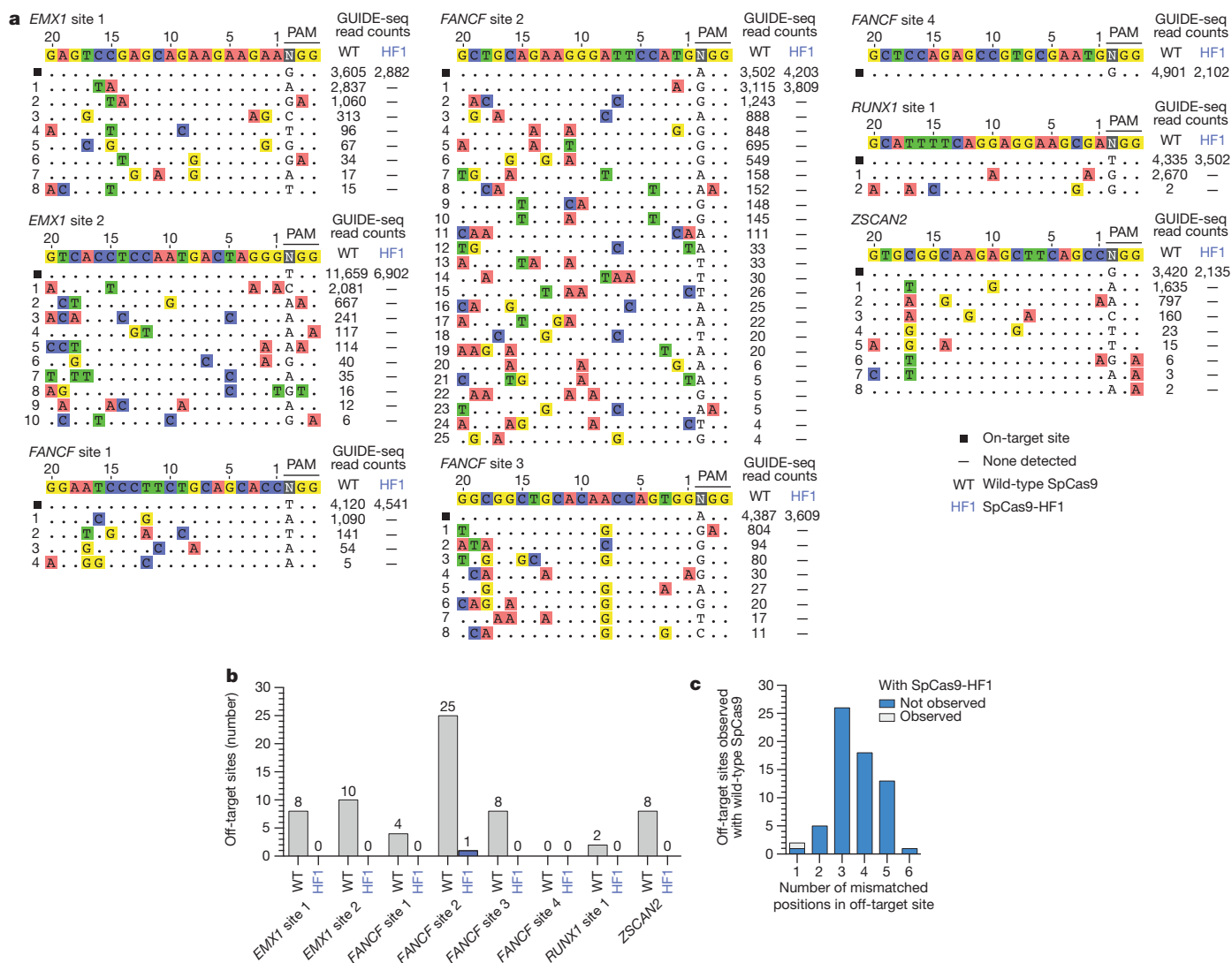


Figure 2 | Genome-wide specificities of wild-type SpCas9 and SpCas9-HF1 with sgRNAs targeted to standard, non-repetitive sites. **a**, Off-target cleavage sites of wild-type SpCas9 and SpCas9-HF1 with eight sgRNAs targeted to endogenous human genes, as determined by GUIDE-seq. Read counts represent a measure of cleavage frequency at a given site; mismatched positions within the spacer or PAM are highlighted

SpCas9 in our GUIDE-seq experiments (four of the 40 sites could not be specifically amplified from genomic DNA). These deep sequencing experiments showed that: (1) wild-type SpCas9 and SpCas9-HF1 induced comparable frequencies of indels at each of the six sgRNA on-target sites, indicating that the nucleases and sgRNAs were functional in all experimental replicates (Fig. 3a, b); (2) as expected, wild-type SpCas9 showed statistically significant evidence of indel mutations at 35 of the 36 off-target sites (Fig. 3b) at frequencies that correlated well with GUIDE-seq read counts for these same sites (Fig. 3c); and (3) the frequencies of indels induced by SpCas9-HF1 at 34 of the 36 off-target sites were statistically indistinguishable from the background level of indels observed in samples from control transfections (Fig. 3b). For the two off-target sites that appeared to have statistically significant mutation frequencies with SpCas9-HF1 relative to the negative control, the mean frequencies of indels were 0.049% and 0.037%, levels at which it is difficult to determine whether these are due to sequencing or PCR error or are bona fide nuclease-induced indels. Based on these results, we conclude that SpCas9-HF1 can completely or nearly completely reduce off-target mutations that occur across a range of different frequencies with wild-type SpCas9 to levels generally undetectable by GUIDE-seq and targeted deep sequencing.

in colour. **b**, Summary of the total number of genome-wide off-target sites identified by GUIDE-seq for wild-type SpCas9 and SpCas9-HF1 with the sgRNAs used in panel **a**. **c**, Off-target sites identified for wild-type SpCas9 and SpCas9-HF1 for the eight sgRNAs, binned according to the total number of mismatches (in the protospacer and PAM) relative to the on-target site.

We next assessed the capability of SpCas9-HF1 to reduce genome-wide off-target effects of sgRNAs designed against atypical homopolymeric or repetitive sequences. Although we and other researchers now try to avoid on-target sites with these characteristics due to their relative lack of orthogonality to the genome, we wished to challenge the genome-wide specificity of SpCas9-HF1 with sites that have very large numbers of known off-target sites in human cells. Therefore, we used previously characterized sgRNAs^{4,8} that target either a cytosine-rich homopolymeric sequence or a sequence containing multiple TG repeats in the human *VEGFA* gene (*VEGFA* site 2 and *VEGFA* site 3, respectively) (Extended Data Table 1). In control experiments, we again found that each of these sgRNAs induced comparable levels of GUIDE-seq dODN tag incorporation (Extended Data Fig. 3c) and indel mutations (Extended Data Fig. 3d) with both wild-type SpCas9 and SpCas9-HF1, demonstrating that SpCas9-HF1 is not impaired in on-target activity with either of these sgRNAs. Importantly, these GUIDE-seq experiments revealed that SpCas9-HF1 was highly effective at reducing off-target sites of these sgRNAs, with 123/144 sites for *VEGFA* site 2 and 31/32 sites for *VEGFA* site 3 not detected (Fig. 4a and Extended Data Fig. 5). Examination of wild-type SpCas9 off-target sites not detected with SpCas9-HF1 showed that they each possessed a

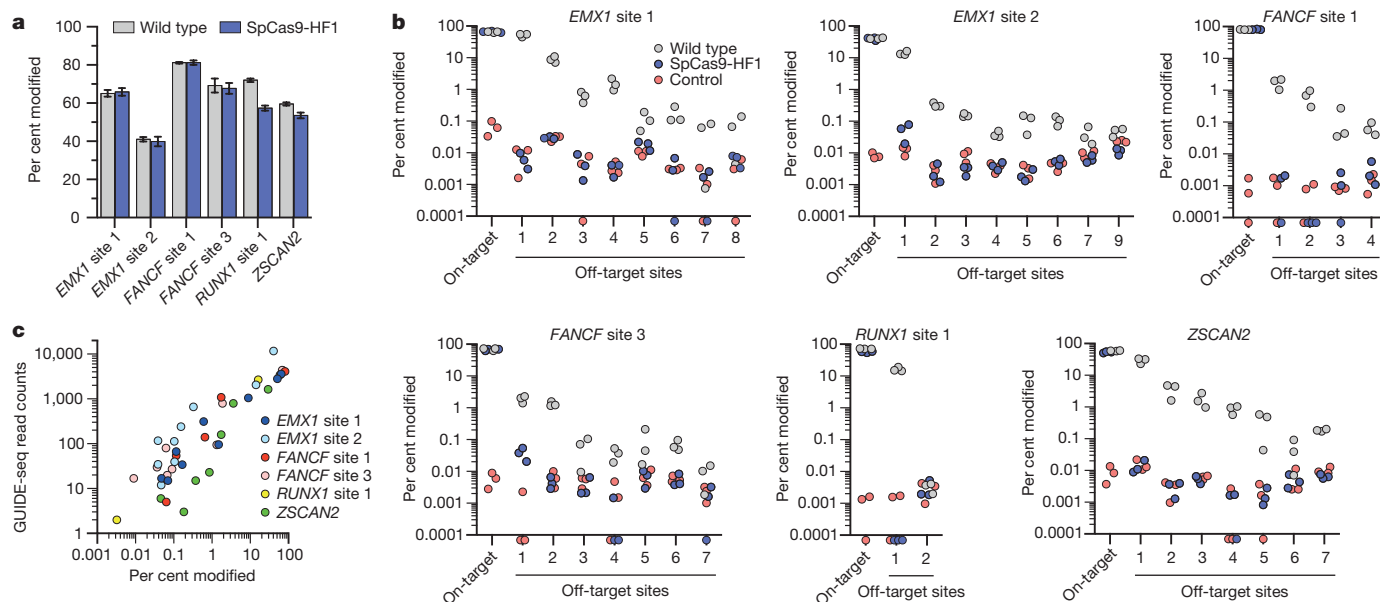


Figure 3 | Validation of SpCas9-HF1 specificity improvements by deep sequencing of off-target sites identified by GUIDE-seq. **a**, Mean on-target per cent modification for wild-type SpCas9 and SpCas9-HF1 with six sgRNAs from Fig. 2. Error bars represent s.e.m. for $n = 3$. **b**, Per cent modification of on-target and GUIDE-seq detected off-target sites with indel mutations. Triplicate experiments are plotted for wild-type SpCas9, SpCas9-HF1, and a negative control; off-target sites are numbered as indicated in Fig. 2a. Filled circles below the x axis represent replicates for which no indels were observed (Supplementary Table 4). Hypothesis testing using a one-sided Fisher exact test with pooled read counts found

significant differences ($P < 0.05$ after adjusting for multiple comparisons using the Benjamini–Hochberg method) for comparisons between SpCas9-HF1 and the control condition only at EMX1 site 2 off-target 1 and FANCF site 3 off-target 1. Significant differences were also found between wild-type SpCas9 and SpCas9-HF1 at all off-target sites, and between wild-type SpCas9 and the control condition at all off-target sites except RUNX1 site 1 off-target 2. **c**, Scatter plot of the correlation between GUIDE-seq read counts (from Fig. 2a) and mean per cent modification determined by deep sequencing at on- and off-target cleavage sites with wild-type SpCas9.

range of total mismatches distributed at various positions within their protospacer and PAM sequences: 2 to 7 mismatches for the VEGFA site 2 sgRNA and 1 to 4 mismatches for the VEGFA site 3 sgRNA (Fig. 4b and Extended Data Fig. 4b); also, nine of these off-targets for VEGFA site 2 may be recognized by an alternate potential base pairing interaction with the sgRNA that might occur with a single bulged base¹² at the sgRNA–DNA interface (Extended Data Figs 5 and 6). Overall, the sites that were still mutated by SpCas9-HF1 possessed a range of 2 to 6 mismatches for the VEGFA site 2 sgRNA and 2 mismatches in the single site for the VEGFA site 3 sgRNA (Fig. 4b), with three of the off-target sites for the VEGFA site 2 sgRNA having an alternative potential single bulge alignment (Extended Data Figs 5 and 6). Notably, no new nuclease-induced off-target sites were induced by SpCas9-HF1 with either of the two sgRNAs. Collectively, these results demonstrate that SpCas9-HF1 can be highly effective at reducing off-target effects of sgRNAs targeted to simple repeat sequences and can also have substantial impacts on sgRNAs targeted to homopolymeric sequences.

Refining the specificity of SpCas9-HF1

Previously described methods such as truncated sgRNAs¹⁴ and the SpCas9 D1135E variant¹⁵ can partially reduce SpCas9 off-target effects, and we therefore wondered whether these might be combined with SpCas9-HF1 to further improve its genome-wide specificity. Testing of SpCas9-HF1 with matched full-length and truncated sgRNAs targeted to four sites in the human cell-based EGFP disruption assay revealed that shortening sgRNA complementarity length substantially impaired on-target activities (Extended Data Fig. 7a). By contrast, SpCas9-HF1 with an additional D1135E substitution (a variant we call SpCas9-HF2) retained 70% or more activity of wild-type SpCas9 with six of eight sgRNAs tested using our human cell-based EGFP disruption assay (Fig. 5a and Extended Data Fig. 2b). We also constructed SpCas9-HF3 and SpCas9-HF4 variants harbouring additional L169A or Y450A substitutions, respectively, at positions whose side chains are believed to mediate non-specific hydrophobic interactions with the target DNA

on its PAM proximal end^{28,31} (Fig. 1a). The Y450 residue is notable for participating in a base stacking interaction with the sgRNA³¹ and undergoing a 120 degree shift upon target binding to create its hydrophobic interaction with the DNA^{28,32}. SpCas9-HF3 and SpCas9-HF4 retained 70% or more of the activities observed with wild-type SpCas9 with the same six out of eight EGFP-targeted sgRNAs (Fig. 5a and Extended Data Fig. 2b).

We next sought to determine whether SpCas9-HF2, -HF3, or -HF4 could reduce indel frequencies at two off-target sites that remained susceptible to modification by SpCas9-HF1, one with the FANCF site 2 sgRNA and another with the VEGFA site 3 sgRNA. For the FANCF site 2 off-target, which bears a single mismatch in the seed sequence of the protospacer, we found that SpCas9-HF4 (containing the additional Y450A substitution) reduced indel mutation frequencies to near background level as judged by T7 endonuclease I assay, while also beneficially

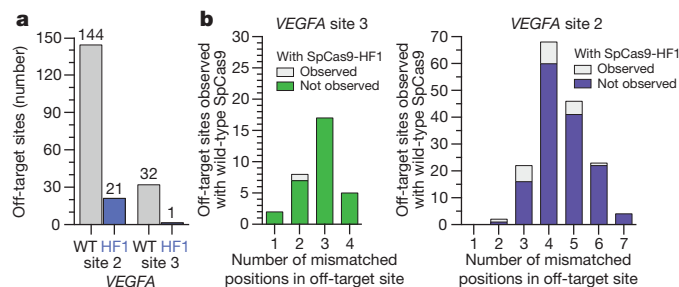


Figure 4 | Genome-wide specificities of wild-type SpCas9 and SpCas9-HF1 with sgRNAs targeted to non-standard, repetitive sites. **a**, Summary of the total number of genome-wide off-target cleavage sites identified by GUIDE-seq for wild-type SpCas9 and SpCas9-HF1 with sgRNAs targeted to VEGFA sites 2 and 3. **b**, Off-target sites identified for wild-type SpCas9 or SpCas9-HF1 with sgRNAs targeted to VEGFA sites 2 and 3 binned according to the total number of mismatches (within the protospacer and PAM) relative to the on-target site.

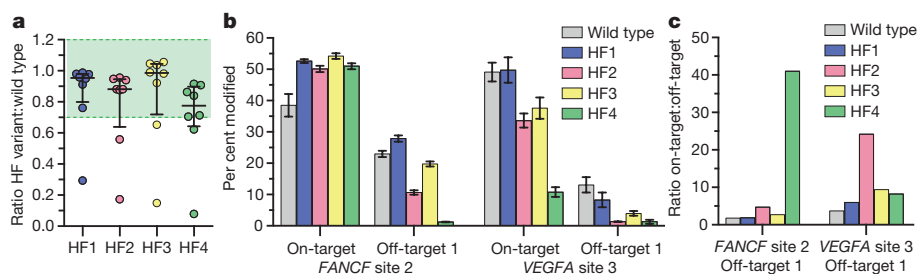


Figure 5 | Activities of high-fidelity derivatives of SpCas9-HF1 bearing additional substitutions. **a**, Summary of the on-target EGFP disruption activities of various SpCas9-HF variants compared to wild-type SpCas9 (from the data in Extended Data Fig. 2b). SpCas9-HF1 contains N497A, R661A, Q695A, and Q926A substitutions; HF2 = HF1 + D1135E; HF3 = HF1 + L169A; HF4 = HF1 + Y450A. The median and interquartile range are shown; the interval showing >70% of wild-type activity is highlighted in green. **b**, Mean per cent modification by SpCas9

increasing on-target activity (Fig. 5b), resulting in the greatest increase in specificity among the three variants (Fig. 5c). For the *VEGFA* site 3 off-target site, which bears two protospacer mismatches (one in the seed sequence and one at the nucleotide most distal from the PAM sequence), SpCas9-HF2 (containing the additional D1135E substitution) showed near background levels of indel formation as determined by T7 endonuclease I assay while showing modest effects on on-target mutation efficiency (Fig. 5b), leading to the greatest increase in specificity for this off-target site among the three variants tested (Fig. 5c).

Discussion

The SpCas9-HF1 variant characterized in this report reduces all or nearly all genome-wide off-target effects to undetectable levels as judged by GUIDE-seq and targeted next-generation sequencing, with the most robust and consistent effects observed with sgRNAs designed against standard, non-repetitive target sequences. Our observations suggest that off-target mutations might be minimized by using SpCas9-HF1 to target non-repetitive sequences that do not have closely matched sites (for example, bearing 1 or 2 mismatches) elsewhere in the genome; such sites can be easily identified using existing publicly available software programs³³. An interesting question will be to determine whether SpCas9-HF1 induces off-target mutations at frequencies below the detection limit of existing unbiased genome-wide methods (Supplementary Discussion). We also discuss other practical considerations for targeting sites of interest with SpCas9-HF1, including the use of sgRNAs with non-G or mismatched 5' nucleotides (Extended Data Fig. 7b) and altering the PAM recognition specificity of SpCas9-HF1 (Extended Data Fig. 8), in the Supplementary Discussion.

Further biochemical experiments and structural characterization will be required to define the mechanism by which SpCas9-HF1 achieves its high genome-wide specificity. We do not believe that the four substitutions we introduced alter the stability or steady-state expression level of SpCas9 in human cells, because titration experiments with decreasing concentrations of expression plasmids suggest that wild-type SpCas9 and SpCas9-HF1 behave comparably as their amounts are lowered (Extended Data Fig. 9). Although our initial rationale for making the substitutions in SpCas9-HF1 was to decrease the energetics of interaction between the Cas9-sgRNA complex and the target DNA (as has been previously proposed to explain the increased specificities of transcription activator-like effector nucleases bearing substitutions at positively charged residues³⁴), recent work has provided greater mechanistic insights into SpCas9 recognition and cleavage. These studies suggest alternative and more detailed models (for example, formation of an active cleavage complex through conformational changes or kinetics of off-target site recognition^{35,36} that might be affected by the substitutions in our SpCas9-HF1 variant (Supplementary Discussion)).

and HF variants at the *FANCF* site 2 and *VEGFA* site 3 on-target sites, as well as off-target sites from Fig. 2a and Extended Data Fig. 5 resistant to the effects of SpCas9-HF1. Per cent modification determined by T7 endonuclease I assay; background indel percentages were subtracted for all experiments; error bars represent s.e.m. for $n = 3$. **c**, Specificity ratios of wild-type SpCas9 and HF variants with the *FANCF* site 2 or *VEGFA* site 3 sgRNAs, plotted as the ratio of on-target to off-target activity (from panel b).

More broadly, our results validate a general strategy for the engineering of additional high-fidelity variants of CRISPR-associated nucleases. We found that introducing substitutions at other non-specific DNA contacting residues can further reduce some of the very small number of residual off-target sites that persist for certain sgRNAs with SpCas9-HF1. Thus, we envision that variants such as SpCas9-HF2, SpCas9-HF4, and others might be used in a customized fashion to eliminate any potential off-target sites that might be resistant to the specificity improvements of SpCas9-HF1. In addition, our variants might be combined with substitutions in residues that contact the non-target DNA strand, alterations that have been shown to reduce SpCas9 off-target effects while our manuscript was under review³⁷. Overall, our results demonstrate that the approach of mutating non-specific DNA contacts is highly effective at increasing SpCas9 specificity and suggest it might be extended to other naturally occurring and engineered Cas9 orthologues^{38–42}, as well as other CRISPR-associated nucleases^{43,44}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 November; accepted 9 December 2015.

Published online 6 January 2016.

- Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
- Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnol.* **32**, 347–355 (2014).
- Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
- Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnol.* **31**, 822–826 (2013).
- Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31**, 827–832 (2013).
- Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnol.* **31**, 839–843 (2013).
- Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β -globin and *CCR5* genes have substantial off-target activity. *Nucleic Acids Res.* **41**, 9584–9592 (2013).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnol.* **33**, 187–197 (2015).
- Frock, R. L. et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnol.* **33**, 179–186 (2015).
- Wang, X. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nature Biotechnol.* **33**, 175–178 (2015).
- Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods* **12**, 237–243 (2015).
- Lin, Y. et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42**, 7473–7485 (2014).
- Cho, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
- Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnol.* **32**, 279–284 (2014).

15. Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas9 nucleases with altered specificities. *Nature* **523**, 481–485 (2015).
16. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnol.* **31**, 833–838 (2013).
17. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
18. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnol.* **32**, 569–576 (2014).
19. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnol.* **32**, 577–582 (2014).
20. Wyvekens, N., Topkar, V. V., Khayter, C., Joung, J. K. & Tsai, S. Q. Dimeric CRISPR RNA-guided FokI–dCas9 nucleases directed by truncated gRNAs for highly specific genome editing. *Hum. Gene Ther.* **26**, 425–431 (2015).
21. Deltcheva, E. *et al.* CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
22. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
23. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nature Biotechnol.* **31**, 233–239 (2013).
24. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
25. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
26. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
27. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
28. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
29. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
30. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnol.* **30**, 460–465 (2012).
31. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
32. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, <http://dx.doi.org/10.1126/science.1247997> (2014).
33. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
34. Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature Methods* **11**, 429–435 (2014).
35. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113 (2015).
36. Knight, S. C. *et al.* Dynamics of CRISPR–Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
37. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* <http://dx.doi.org/10.1126/science.aad5227> (2015).
38. Ran, F. A. *et al.* *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
39. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).
40. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110**, 15644–15649 (2013).
41. Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR–Cas systems. *Nucleic Acids Res.* **42**, 2577–2590 (2014).
42. Kleinstiver, B. P. *et al.* Broadening the targeting range of *Staphylococcus aureus* CRISPR–Cas9 by modifying PAM recognition. *Nature Biotechnol.* **33**, 1293–1298 (2015).
43. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
44. Shmakov, S. *et al.* Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell* **60**, 385–397 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements B.P.K. is supported by a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. V.P. was supported by the Massachusetts General Hospital (MGH) Department of Pathology. S.Q.T. is supported by an MGH Tosteson and Fund for Medical Discovery Fellowship. J.K.J. is supported by a US National Institutes of Health (NIH) Director's Pioneer Award (DP1 GM105378), NIH R01 GM107427, NIH R01 GM088040, and the Jim and Ann Orr MGH Research Scholar Award.

Author Contributions B.P.K., V.P., and J.K.J. conceived of and designed experiments. B.P.K., V.P., and M.S.P. performed all experiments. N.T.N. contributed to GUIDE-seq library preparation. B.P.K., V.P., M.S.P., S.Q.T., and Z.Z. analysed the data. B.P.K., V.P., and J.K.J. wrote the manuscript with input from all the authors.

Author Information Plasmids encoding the high-fidelity SpCas9, VQR, and VRQR variants described in this manuscript have been deposited with the non-profit plasmid distribution service Addgene (<http://www.addgene.org/crispr-cas>). All sequencing data from this study is available through the NCBI Sequence Read Archive (SRA) under accession number SRP066862. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.K.J. (jjoung@mg.harvard.edu).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Plasmids and oligonucleotides. DNA sequences of plasmids used in this study can be found in the Supplementary Information. sgRNA target sites are available in Supplementary Table 1, and oligonucleotides used in this study can be found in Supplementary Table 2. SpCas9 expression plasmids containing amino acid substitutions were generated by standard PCR and molecular cloning into JDS246 (ref. 4). sgRNA expression plasmids were constructed by ligating oligonucleotide duplexes into BsmBI cut BPK1520 (ref. 15). Unless otherwise indicated, all sgRNAs were designed to target sites containing a 5' guanine nucleotide.

Human cell culture and transfection. U2OS cells (a gift from Toni Cathomen, Freiburg) and U2OS.EGFP cells (containing a single integrated copy of a reporter gene encoding an EGFP-PEST fusion)³⁰ were cultured in advanced DMEM supplemented with 10% heat-inactivated fetal bovine serum, 2 mM GlutaMax, and penicillin and streptomycin at 37 °C with 5% CO₂. The growth media for U2OS.EGFP cells was additionally supplemented with 400 µg ml⁻¹ Geneticin. All cell culture reagents were obtained from Life Technologies. Cell line identity was validated by STR profiling (ATCC) and deep-sequencing, and cells were tested bi-weekly for mycoplasma contamination. Unless otherwise noted, cells were co-transfected with 750 ng of Cas9 plasmid and 250 ng of sgRNA plasmid. For negative control experiments, Cas9 plasmids were co-transfected with a U6-null plasmid. Nucleofections were performed using the DN-100 program on a Lonza 4-D Nucleofector with the SE Cell Line Kit according to the manufacturer's protocol (Lonza). For T7 endonuclease I assays, GUIDE-seq experiments, and targeted deep sequencing, genomic DNA was extracted ~72 h post-transfection using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter Genomics).

Human cell EGFP disruption assay. EGFP disruption experiments, in which cleavage and induction of indels by non-homologous end-joining (NHEJ)-mediated repair within a single integrated EGFP reporter gene leads to loss of cell fluorescence, were performed as previously described^{4,30}. Briefly, transfected cells were analysed ~52 h post-transfection for loss of EGFP expression using a Fortessa flow cytometer (BD Biosciences). Background EGFP loss was determined using negative control transfections gated at ~2.5% for all experiments (represented as a red dashed line in figures). *P* values for comparisons between SpCas9 variants were calculated using a one-sided *t*-test with equal variances and adjusted for multiple comparisons using the method of Benjamini and Hochberg (Supplementary Table 3).

T7 endonuclease I assays. To quantify mutagenesis frequencies at desired genomic loci, T7 endonuclease I assays were performed as previously described³⁰. Briefly, on- or off-target sites were amplified from ~100 ng of genomic DNA using Phusion Hot-Start Flex DNA Polymerase (New England Biolabs) using the primers listed in Supplementary Table 2. An Agencourt Ampure XP cleanup (Beckman Coulter Genomics) was performed before the denaturation and annealing of ~200 ng of the PCR product, followed by digestion with T7 endonuclease I (New England Biolabs). Purified digestion products were quantified using a QIAxcel capillary electrophoresis instrument (Qiagen) to approximate the mutagenesis frequencies induced by Cas9-sgRNA complexes. *P* values for comparisons between SpCas9 variants were calculated using a one-sided *t*-test with equal variances and adjusted for multiple comparisons using the method of Benjamini and Hochberg (Supplementary Table 3).

GUIDE-seq. GUIDE-seq relies on the integration of a short dsODN tag into DNA breaks to enable amplification and sequencing of adjacent genomic sequence, with the number of tag integrations at any given site providing a quantitative measure of cleavage efficiency⁸. GUIDE-seq experiments were performed and analysed essentially as previously described⁸. Briefly, U2OS cells were transfected with 750 ng of Cas9 and 250 ng sgRNA plasmids as described above, along with 100 pmol of a GUIDE-seq end-protected dsODN that contains an NdeI restriction site⁸. Restriction-fragment length polymorphism (RFLP) assays were used to estimate GUIDE-seq tag integration frequencies at the intended on-target sites as previously described¹⁵, using the primers listed in Supplementary Table 2.

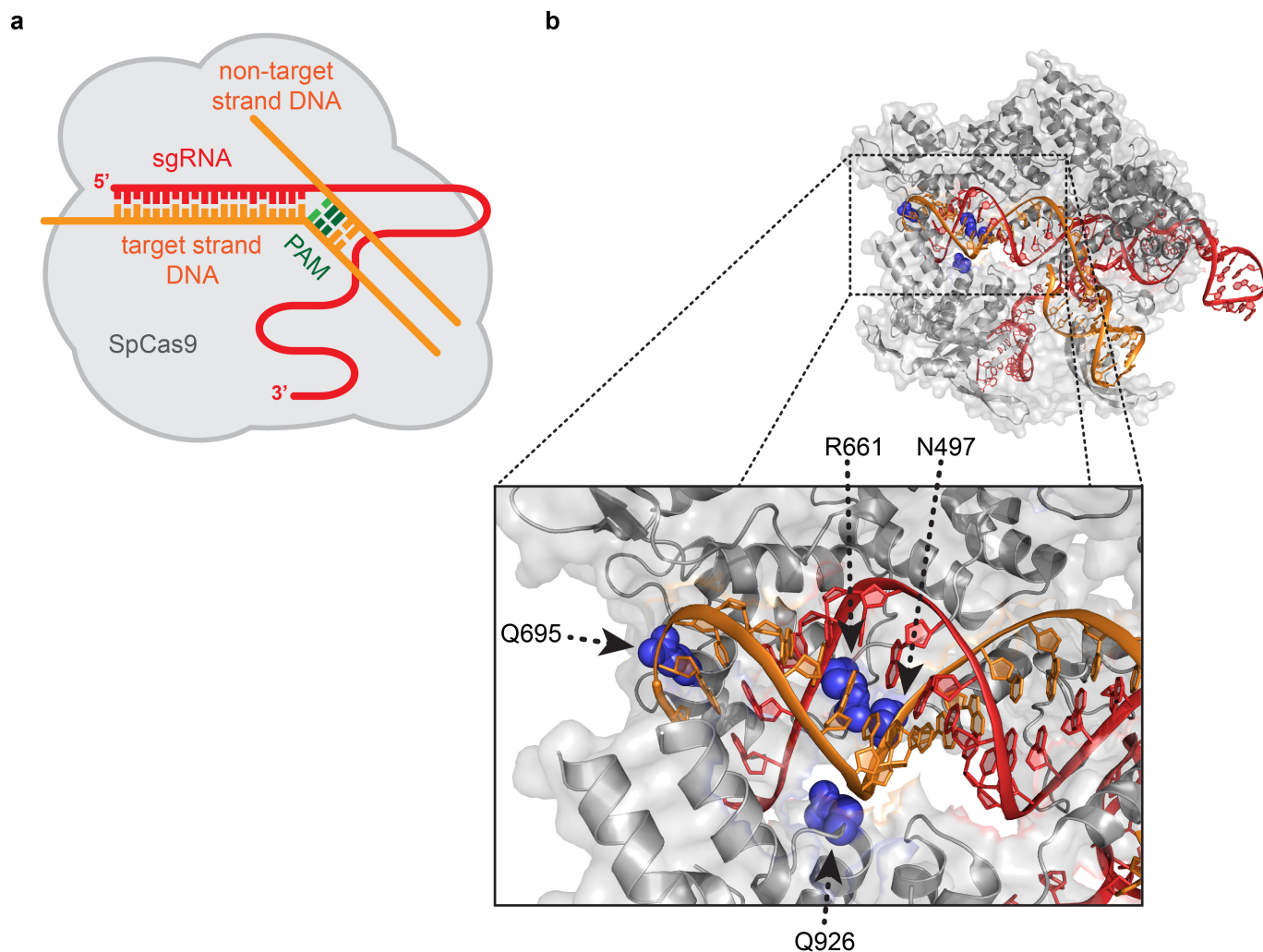
The overall on-target mutagenesis frequencies of GUIDE-seq tag-treated samples was determined by T7 endonuclease I assay as described above. Tag-specific amplification and library preparation⁸ were performed before high-throughput sequencing on an Illumina MiSeq instrument. GUIDE-seq data was analysed as previously described⁸ using open-source GUIDE-seq analysis software (<http://www.jounglab.org/guideseq>) and the summarized results can be found in Supplementary Table 4. Genomic sites were excluded from analysis on the basis of overlap with background genomic breakpoint regions detected in any of four oligo-only control samples, overlap with previously identified Cas9-sgRNA independent breakpoints in human U2OS cells⁸, or as neighbouring genomic window consolidation artefacts likely due to extensive end-resection around breakpoints (Supplementary Table 4). Potential RNA- or DNA-bulge sites¹² (Extended Data Fig. 6) were identified by sequence alignment with Geneious version 8.1.6 (<http://www.geneious.com>)⁴⁵. Sequencing data was corrected for U2OS cell-type specific SNPs with the site encoding the smallest edit distance to the intended sgRNA site used as the most likely off-target (Supplementary Table 4). Differences in number of GUIDE-seq identified off-target sites between this work and previous studies^{8,15} are likely due to different experimental conditions (for example, different promoters, quantity of plasmids used for transfection) and/or to sampling effects at the limit of detection of these particular experiments (Supplementary Table 4), and most likely not due to depth of sequencing which was similar between experiments.

Positional profiles generated from GUIDE-seq data (Extended Data Fig. 4) were made by weighting each nucleotide at each on/off-target site by the number of GUIDE-seq read counts. Sites containing gapped alignments relative to the human genome were not considered. Positional profiles for potential genomic off-target sites were restricted to sequences containing five or fewer mutations relative to the on-target site and to sequences containing NGG PAMs. Heat maps were generated with R 3.2.2 and the image function, with colours determined using the function `colorRampPalette(c("white", "blue"))(2500)`.

Targeted deep-sequencing. Off-target sites identified by GUIDE-seq were amplified using Phusion High-Fidelity DNA polymerase (New England Biolabs) using the primers listed in Supplementary Table 2 for the genomic amplicons listed in Supplementary Table 5. PCR products were generated for each on- and off-target site from ~100 ng of genomic DNA extracted from U2OS cells. Products were generated from triplicate transfections for each of three experimental conditions: (1) control (wild-type SpCas9 + pSL695, a control plasmid that contains a U6 promoter but does not encode a functional sgRNA), (2) wild-type SpCas9 + sgRNA, and (3) SpCas9-HF1 + sgRNA. PCR products were purified with Ampure XP magnetic beads (Agencourt), normalized in concentration, and pooled into nine samples (individual triplicate experiments for each of the three conditions listed above). Illumina Tru-seq compatible deep-sequencing libraries were prepared using ~500 ng of each pooled sample using a 'with-bead' HTP library preparation kit (KAPA BioSystems), and sequenced via 150-bp paired-end sequencing on an Illumina MiSeq instrument. High-throughput sequencing data was analysed essentially as previously described¹⁸. Briefly, paired reads were mapped to the human genome (reference sequence GRCh37) using the bwa mem algorithm with default parameters. High-quality reads (average quality score ≥30) were analysed for the presence of two or more bp indels that overlapped to the on- or off-target sites (Supplementary Table 5). One bp indel mutations were only included if they occurred directly adjacent to the predicted cleavage site. *P* values for comparisons between control, wild-type SpCas9 + sgRNA, and SpCas9-HF1 + sgRNA (Supplementary Table 5) were obtained on pooled triplicate data using a one-sided Fisher exact test in the R 3.2.2 software package. *P* values for each set of comparisons were adjusted for multiple comparisons using the method of Benjamini and Hochberg (function `p.adjust(method = "BH")` in R).

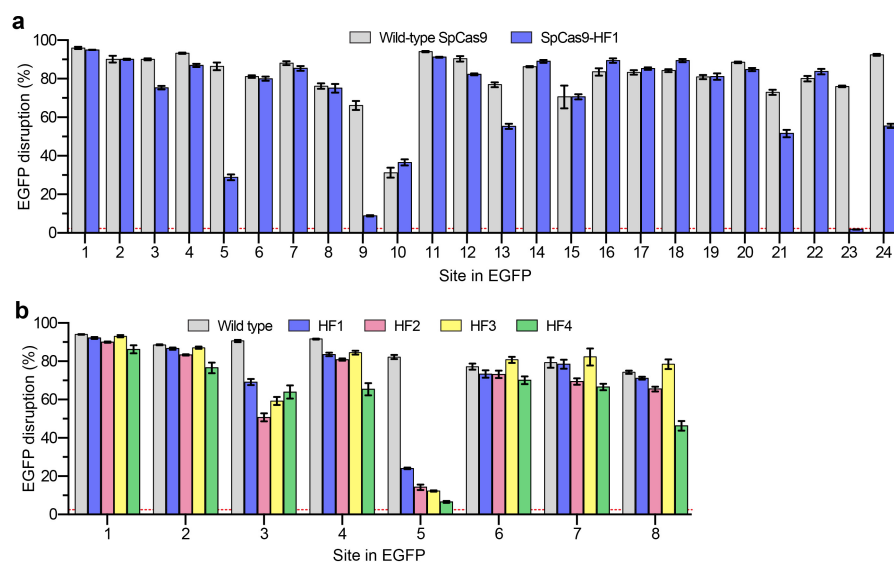
Code availability. Scripts for GUIDE-seq analysis (v0.9) can be found at <http://jounglab.org/guideseq>. The scripts used for indel calling on deep sequencing data and GUIDE-seq profiles are available upon request.

45. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).



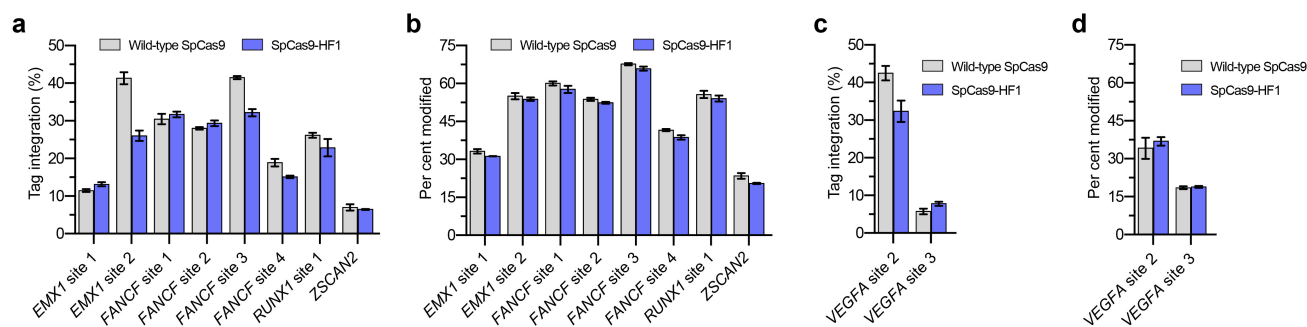
Extended Data Figure 1 | SpCas9 interaction with the sgRNA and target DNA. **a**, Schematic illustrating the SpCas9-sgRNA complex, with base pairing between the sgRNA and target DNA. **b**, Structural representation of the SpCas9-sgRNA complex bound to the target DNA, from PDB

accession code 4UN3 (ref. 29). The four residues that form hydrogen bond contacts to the target-strand DNA backbone are highlighted in blue, the HNH domain is hidden for visualization purposes.



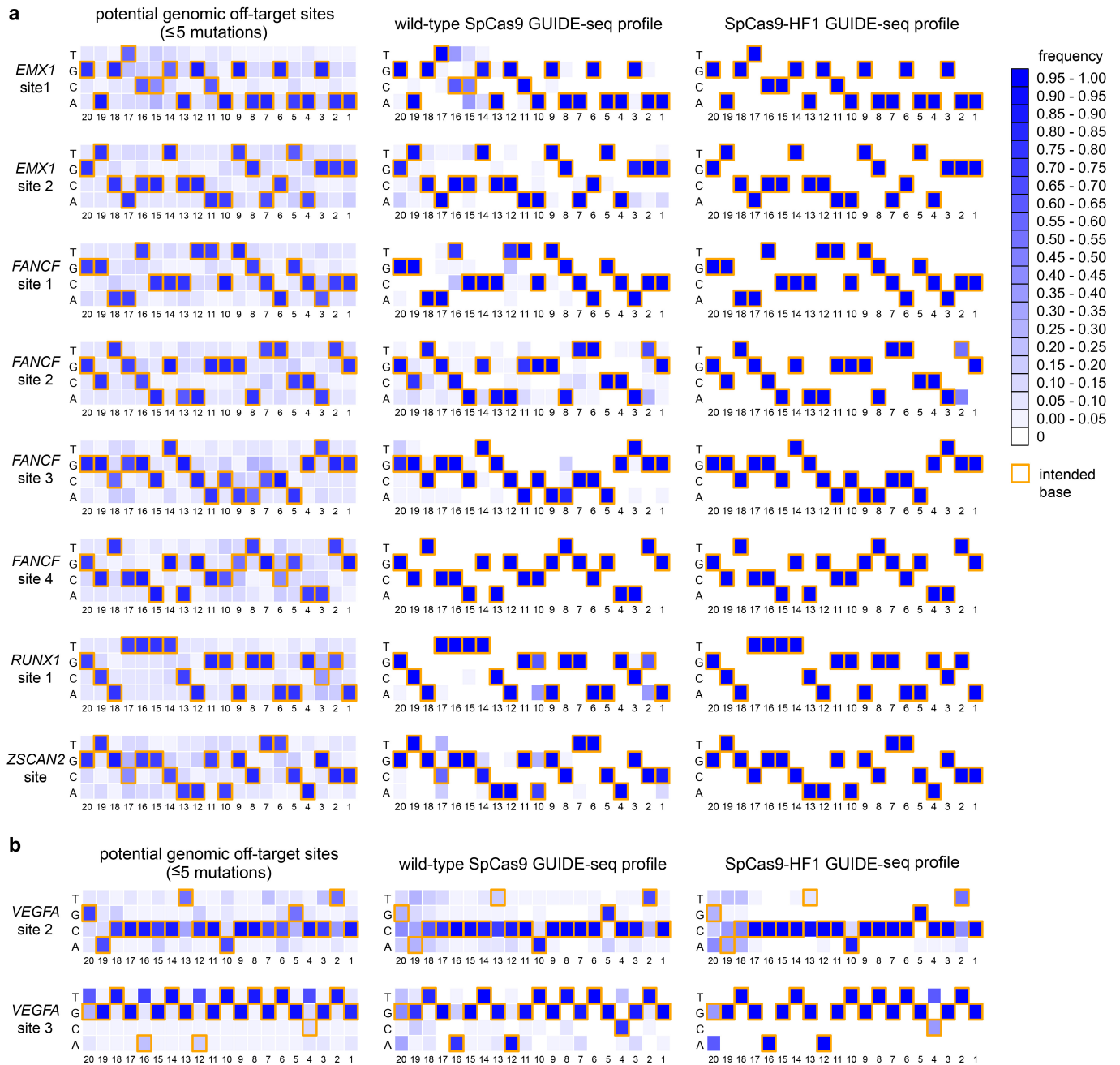
Extended Data Figure 2 | On-target activities of high-fidelity SpCas9 variants. **a, b,** EGFP disruption activities of wild-type SpCas9 and SpCas9-HF1 (**a**) and SpCas9-HF1-derivative variants (**b**) in human cells. SpCas9-HF1 contains N497A, R661A, Q695A, and Q926A substitutions;

HF2 = HF1 + D1135E; HF3 = HF1 + L169A; HF4 = HF1 + Y450A. Error bars represent s.e.m. for $n = 3$; mean level of background EGFP loss represented by the red dashed line.



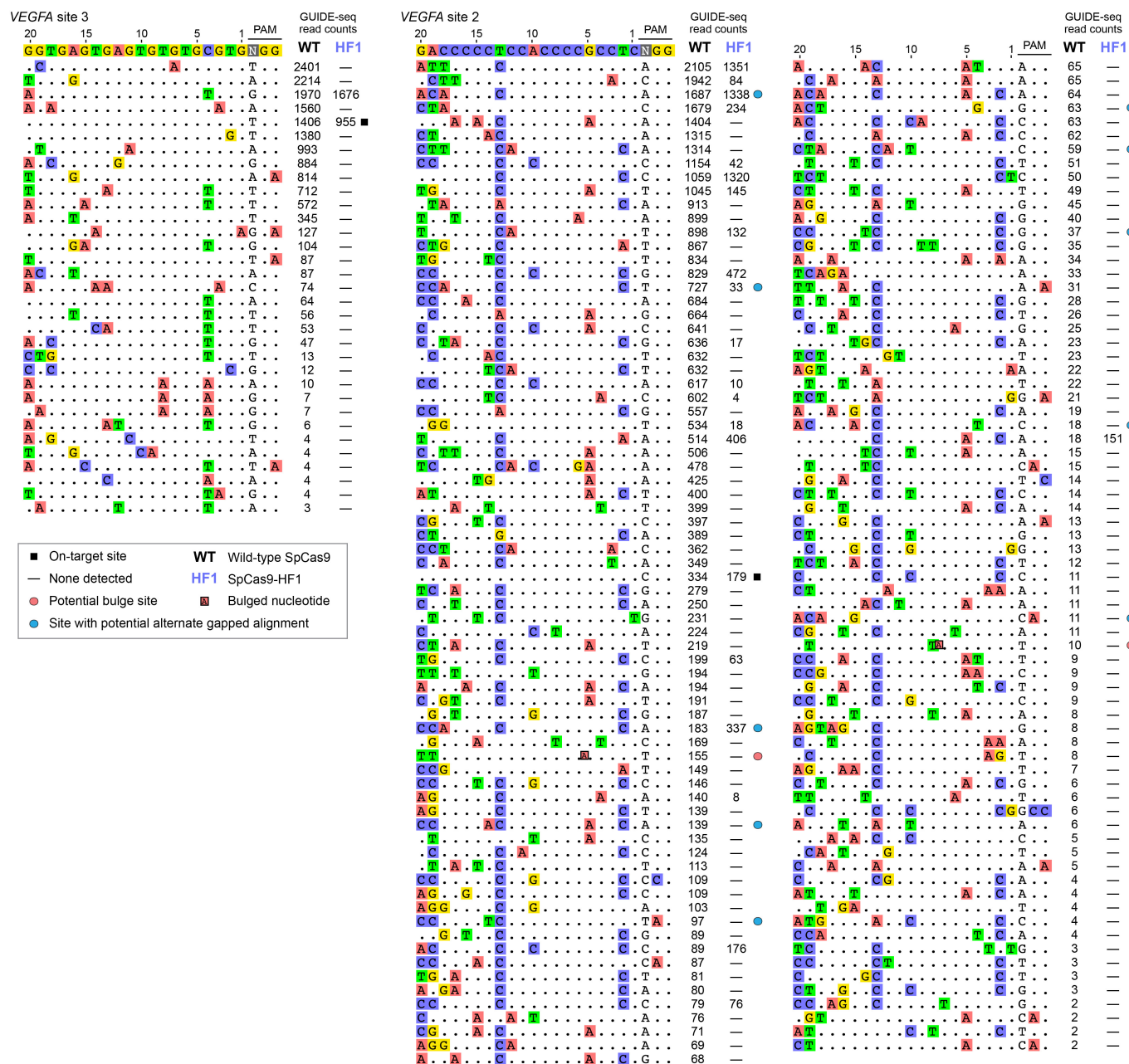
Extended Data Figure 3 | On-target activity comparisons of wild-type and SpCas9-HF1 with various sgRNAs used for GUIDE-seq experiments. a, c, Mean GUIDE-seq tag integration at the intended on-target site for GUIDE-seq experiments shown in Figs 2a and Extended Data Fig. 5 (a and c, respectively), quantified by restriction-fragment

length polymorphism assay. Error bars represent s.e.m. for $n = 3$. **b, d,** Mean percent modification at the intended on-target site for GUIDE-seq experiments shown in Fig. 2a and Extended Data Fig. 5 (b and d, respectively), detected by T7 endonuclease I assay. Error bars represent s.e.m. for $n = 3$.



Extended Data Figure 4 | Positional summary of off-target sites identified by GUIDE-seq. a, b, Heat maps derived from GUIDE-seq data with sgRNAs targeting non-repetitive (**a**), or repetitive or homopolymeric sites (**b**) in the genome are shown. Base frequencies in the set of all potential genomic off-target sites (weighted equally) with NGG PAMs and five or fewer mutations for each sgRNA are shown on the left. Summaries of off-target sites identified by GUIDE-seq for wild-type SpCas9 and SpCas9-HF1 (both weighted by read count) are shown on

the right. Yellow box outlines denote on-target bases at each position. Positions (20–1) are shown below the heat maps, with 1 being the most PAM-proximal position. Note the presence of mismatches that would be expected to create potential wobble interactions (G→A or T→C) at certain positions among the off-target sites induced by wild-type SpCas9 and that SpCas9-HF1 appears to reduce off-target activity without any obvious positional bias.

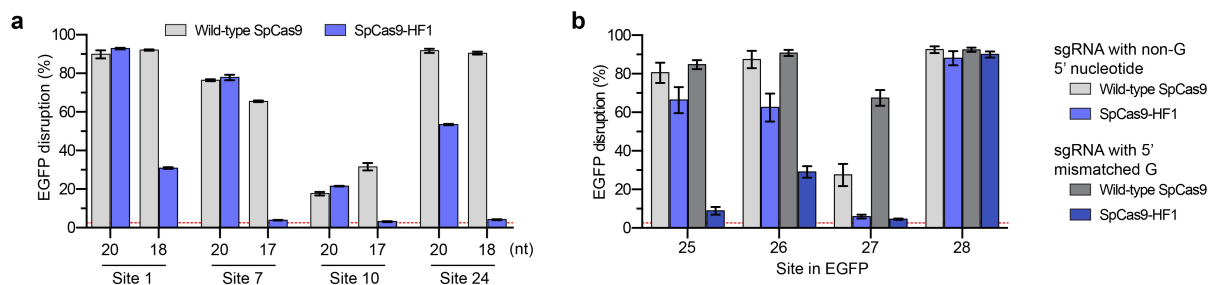


Extended Data Figure 5 | Genome-wide cleavage specificity of wild-type SpCas9 and SpCas9-HF1 with sgRNAs targeted to non-standard, repetitive sites. a, GUIDE-seq profiles of wild-type SpCas9 and SpCas9-HF1 using two sgRNAs known to cleave large numbers of off-target sites^{4,8}. GUIDE-seq read counts represent a measure of cleavage efficiency at a given site. Mismatched positions within the spacer or PAM are highlighted in colour red circles indicate off-target sites likely to have

the indicated bulge¹² at the sgRNA–DNA interface, blue circles indicate sites that may have an alternative gapped alignment relative to the one shown (see Extended Data Fig. 6). Off-target sites marked with red circles are not included in the counts of Fig. 4b, sites marked with blue circles are counted with the number of mismatches in the non-gapped alignment for Fig. 4b.

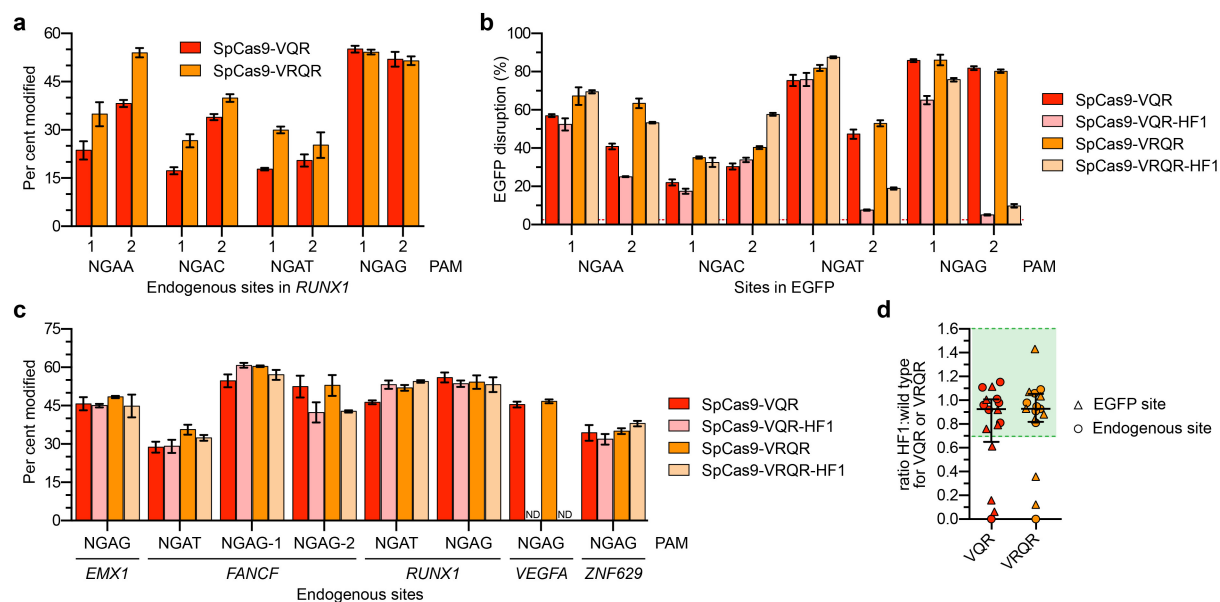
	Traditional alignment	Alignment containing gaps
VEGFA site 2 off-target #3	GACCCCTCCACCCGCGCTCNGG ACA...C...A..	GACCCCTCCACCCGCGCTCNGG C...-...A..
VEGFA site 2 off-target #17	GACCCCTCCACCCGCGCTCNGG CCA...C...C.T..	GACCCCTCCACCCGCGCTCNGG C...-...C.T..
VEGFA site 2 off-target #49	GACCCCTCCACCCGCGCTCNGG CCA...C...C.A..	GACCCCTCCACCCGCGCTCNGG C...-...C.A..
VEGFA site 2 off-target #56	GACCCCTCCACCCGCGCTCNGG CC...AC...A.C.A..	GACCCCTCCACCCGCGCTCNGG C...AC...A.C.A..
VEGFA site 2 off-target #63	GACCCCTCCACCCGCGCTCNGG CC...TC...TA..	GACCCCTCCACCCGCGCTCNGG C...C...TA..
VEGFA site 2 off-target #77	GACCCCTCCACCCGCGCTCNGG AC...G...G..	GACCCCTCCACCCGCGCTCNGG C...C...G...G..
VEGFA site 2 off-target #80	GACCCCTCCACCCGCGCTCNGG CTA...CA...C..	GACCCCTCCACCCGCGCTCNGG C...A...C..
VEGFA site 2 off-target #86	GACCCCTCCACCCGCGCTCNGG CC...TC...C.G..	GACCCCTCCACCCGCGCTCNGG CC...C...C.G..
VEGFA site 2 off-target #100	GACCCCTCCACCCGCGCTCNGG AC...A.C...C..	GACCCCTCCACCCGCGCTCNGG C...A.C...C..
VEGFA site 2 off-target #114	GACCCCTCCACCCGCGCTCNGG ACA...G...CA..	GACCCCTCCACCCGCGCTCNGG A...A...G...CA..

Extended Data Figure 6 | Potential alternate alignments for VEGFA site 2 off-target sites. Ten VEGFA site 2 off-target sites identified by GUIDE-seq (left) that may potentially be recognized as off-target sites with single nucleotide gaps¹² (right), aligned using Geneious⁴⁵ version 8.1.6 (<http://www.geneious.com>).



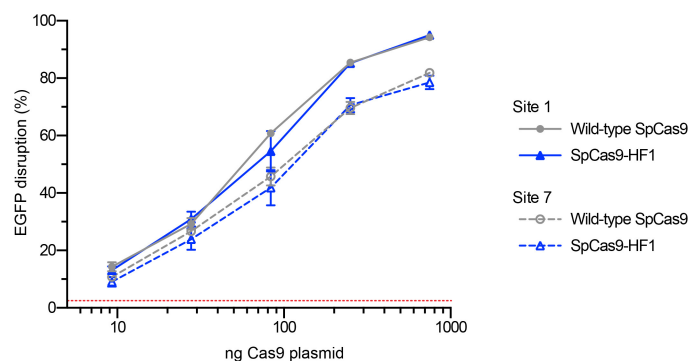
Extended Data Figure 7 | Activities of wild-type SpCas9 and SpCas9-HF1 with truncated or 5' mismatched sgRNAs¹⁴. **a**, EGFP disruption activities of wild-type SpCas9 and SpCas9-HF1 using full-length or truncated sgRNAs. **b**, EGFP disruption activities of wild-type SpCas9 and SpCas9-HF1 using sgRNAs that encode a matched

5' non-G nucleotide or an intentionally mismatched 5' G nucleotide. For both panels, error bars represent s.e.m. for $n = 3$, and the mean level of background EGFP loss observed in control experiments is represented by the red dashed line.



Extended Data Figure 8 | Altering the PAM recognition specificity of SpCas9-HF1. **a**, Comparison of the mean per cent modification of on-target endogenous human sites by the SpCas9-VQR variant (ref. 15) and an improved SpCas9-VRQR variant using 8 sgRNAs, quantified by T7 endonuclease I assay. Both variants are engineered to recognize an NGAN PAM. Error bars represent s.e.m. for $n = 3$. **b**, On-target EGFP disruption activities of SpCas9-VQR and SpCas9-VRQR compared to their -HF1 counterparts using eight sgRNAs. Error bars represent s.e.m. for $n = 3$; mean level of background EGFP loss in negative controls

represented by the red dashed line. **c**, Comparison of the mean on-target per cent modification by SpCas9-VQR and SpCas9-VRQR compared to their -HF1 variants at eight endogenous human gene sites, quantified by T7 endonuclease I assay. Error bars represent s.e.m. for $n = 3$; ND, not detectable. **d**, Summary of the fold-change in on-target activity when using SpCas9-VQR or SpCas9-VRQR compared to their corresponding -HF1 variants (from **b** and **c**). The median and interquartile range are shown, the interval showing greater than 70% of wild-type activity is highlighted in green.



Extended Data Figure 9 | Titrations of wild-type SpCas9 and SpCas9-HF1 expression plasmid amounts. Human cell EGFP disruption activities from transfections with varying amounts of wild-type and SpCas9-HF1 expression plasmids. For all transfections, the amount of sgRNA-containing plasmid was fixed at 250 ng. Two sgRNAs targeting different sites were used; Error bars represent s.e.m. for $n = 3$; mean level of background EGFP loss in negative controls is represented by the red dashed line.

Extended Data Table 1 | Summary of potential mismatched sites in the reference human genome for the ten sgRNAs examined by GUIDE-seq

site	spacer with PAM	mismatches to on-target site*						total
		1	2	3	4	5	6	
<i>EMX1</i> site 1	GAGTCCGAGCAGAAGAAGAAGGG	0	1	18	273	2318	15831	18441
<i>EMX1</i> site 2	GTCACCTCCAATGACTAGGGTGG	0	0	3	68	780	6102	6953
<i>FANCF</i> site 1	GGAATCCCTTCTGCAGCACCTGG	0	1	18	288	1475	9611	11393
<i>FANCF</i> site 2	GCTGCAGAAGGGATTCCATGAGG	1	1	29	235	2000	13047	15313
<i>FANCF</i> site 3	GGCGGCTGCACAACCACTGGAGG	0	0	11	79	874	6651	7615
<i>FANCF</i> site 4	GCTCCAGAGCCGTGCGAATGGGG	0	0	6	59	639	5078	5782
<i>RUNX1</i> site 1	GCATTTTCAGGAGGAAGCGATGG	0	2	6	189	1644	11546	13387
<i>ZSCAN2</i>	GTGCGGCAAGAGCTTCAGCCGGG	0	3	12	127	1146	10687	11975
<i>VEGFA</i> site 2	GACCCCTCCACCCCGCTCCGG	0	2	35	456	3905	17576	21974
<i>VEGFA</i> site 3	GGTGAGTGAGTGTGTGCGTGTGG	1	17	383	6089	13536	35901	55927

*Determined using Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>).

Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions

Alexander J. Westermann¹, Konrad U. Förstner^{1,2}, Fabian Amman^{3,4}, Lars Barquist¹, Yanjie Chao¹, Leon N. Schulte¹, Lydia Müller³, Richard Reinhardt⁵, Peter F. Stadler^{3,4,6,7} & Jörg Vogel^{1,8}

Bacteria express many small RNAs for which the regulatory roles in pathogenesis have remained poorly understood due to a paucity of robust phenotypes in standard virulence assays. Here we use a generic ‘dual RNA-seq’ approach to profile RNA expression simultaneously in pathogen and host during *Salmonella enterica* serovar Typhimurium infection and reveal the molecular impact of bacterial riboregulators. We identify a PhoP-activated small RNA, PinT, which upon bacterial internalization temporally controls the expression of both invasion-associated effectors and virulence genes required for intracellular survival. This riboregulatory activity causes pervasive changes in coding and noncoding transcripts of the host. Interspecies correlation analysis links PinT to host cell JAK–STAT signalling, and we identify infection-specific alterations in multiple long noncoding RNAs. Our study provides a paradigm for a sensitive RNA-based analysis of intracellular bacterial pathogens and their hosts without physical separation, as well as a new discovery route for hidden functions of pathogen genes.

Regulatory RNAs crucially contribute to post-transcriptional control of gene expression in a wide array of organisms, including pathogenic bacteria^{1–3}. The facultative intracellular pathogen *Salmonella enterica* serovar Typhimurium (hereafter referred to as *Salmonella*) expresses hundreds of small regulatory RNAs (sRNAs), many of which are activated under defined stress and virulence conditions^{4–10}, suggesting a role during host infection. Likewise, genetic inactivation of Hfq, a protein required by many sRNAs for target mRNA regulation, attenuates *Salmonella* virulence¹¹. However, deletion of sRNA genes typically results in mild, if any, phenotypes in animal models^{12–14}, probably because most sRNAs act to fine-tune gene expression^{2,3}. Therefore, more sensitive approaches are needed to uncover their molecular functions during infection.

RNA-seq provides a sensitive method for global gene expression analysis in infection biology^{15,16}. However, as bacterial infections of eukaryotic cells involve two interacting organisms with profoundly different transcriptomes, RNA-seq studies are commonly restricted to either the pathogen or host after their physical separation¹⁵. Furthermore, they typically focus on messenger RNAs (mRNAs) as a proxy for protein expression, neglecting the vast RNA output from noncoding regions. Theoretically, RNA-seq should permit a simultaneous profiling of all RNA classes in both intracellular bacteria and the eukaryotic host, despite a striking excess of eukaryotic over bacterial RNA^{15,17}. Such a one-step dual RNA-seq¹⁵ analysis separates transcripts *in silico*, rendering the tedious and error-prone physical separation of pathogen and host superfluous. Here, dual RNA-seq has been used to discover how *Salmonella* sRNAs fine-tune gene expression in intracellular bacteria, with widespread consequences for the human host response.

Dual RNA-seq of *Salmonella*-infected human cells

We established dual RNA-seq for *Salmonella* using HeLa cells in which bacterial invasion is followed by prolonged intracellular replication and a well-defined host response¹⁸ with low TLR activity¹⁹, which helps to

reveal host signalling events directly modulated by the pathogen. We built on previous work²⁰ combining green fluorescent protein (GFP)-expressing *Salmonella* and fluorescence-activated cell sorting (FACS) to select host cells containing on average 10 or 75 bacteria at 4 or 24 h post-infection (p.i.), respectively, and also non-invaded control cells (Fig. 1a and Extended Data Fig. 1a–d). Optimized fixation conditions preserved both transcriptomes, as well as the GFP signal during the extended FACS procedure, while also sterilizing the sample (Extended Data Fig. 1e–h).

To assess coverage of pathogen and host transcripts, we first sequenced total RNA without further depletion or enrichment of certain RNA classes²¹, using Illumina technology. Of ~25 million RNA reads obtained from infected cells at 4 h, 98.2% were of human origin and 1.4% of *Salmonella* origin (Fig. 1b), confirming the predicted excess of eukaryotic over bacterial RNA¹⁵. Bacterial reads increased over time, reflecting intracellular replication of the pathogen (Fig. 1b; compare 4 and 24 h GFP⁺). Reads from *Salmonella* or HeLa cells alone mapped to their respective genomes with high stringency (Fig. 1b) and individual transcript levels were highly reproducible amongst biological triplicates (Extended Data Fig. 1i).

Collective sequencing of total RNA captured all major bacterial and eukaryotic transcript classes (Fig. 1b). Stable housekeeping ribosomal RNA (rRNA) transcripts were abundant in both transcriptomes, as were reads corresponding to transfer RNA (tRNA) in the *Salmonella* and stable small nucleolar RNAs (snRNAs, snoRNAs) in the human transcriptome. Other noncoding bacterial RNA classes, that is, sRNAs and antisense transcripts, contributed ~8% of the *Salmonella* reads; the abundance of human regulatory noncoding transcripts ranged from 0.1% for the ~22 nucleotide-long microRNAs (miRNAs) to 15% for the 200 to 100,000 nucleotide-spanning long noncoding RNAs (lncRNAs).

Messenger RNAs, constituting 16–20% of all reads, predict differentially expressed proteins in both organisms during infection, which

¹University of Würzburg, RNA Biology Group, Institute for Molecular Infection Biology, Josef-Schneider-Straße 2/D15, D-97080 Würzburg, Germany. ²University of Würzburg, Core Unit Systems Medicine, Josef-Schneider-Straße 2/D15, D-97080 Würzburg, Germany. ³University of Leipzig, Department of Computer Science and Interdisciplinary Center for Bioinformatics, Härtelstraße 16-18, D-04107 Leipzig, Germany. ⁴University of Vienna, Theoretical Biochemistry Group, Institute for Theoretical Chemistry, Währinger Straße 17, A-1090 Vienna, Austria. ⁵Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, D-50829 Cologne, Germany. ⁶Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany. ⁷Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, New Mexico 87501, USA. ⁸Research Centre for Infectious Diseases (ZINF), University of Würzburg, D-97070 Würzburg, Germany.

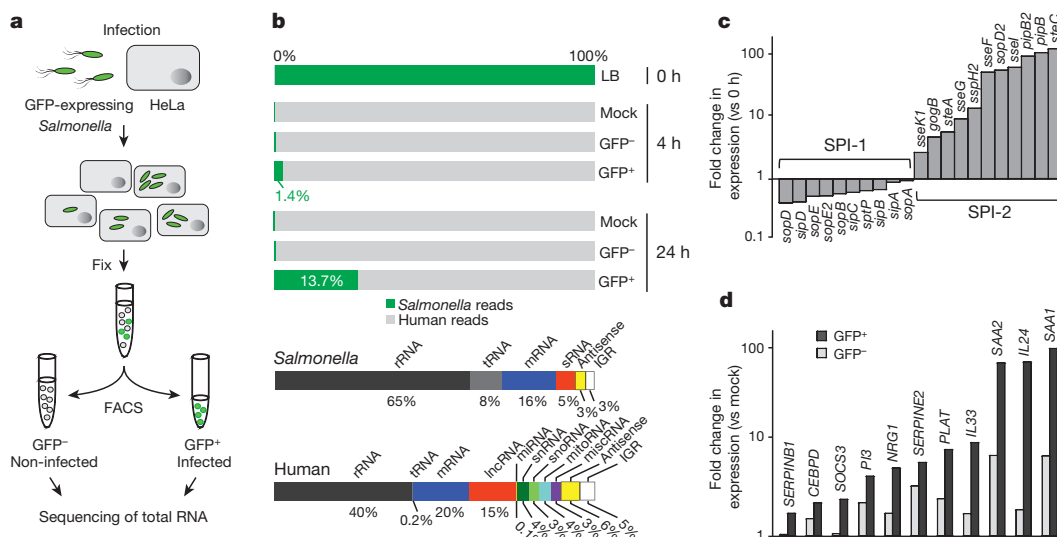


Figure 1 | Dual RNA-seq captures the full transcript repertoire of infected cells. a, Dual RNA-seq workflow. **b**, Representative mapping statistics from sorted *Salmonella*-infected HeLa-S3 cells. LB, *Salmonella* grown in medium. 'miRNA' includes primary and mature forms; 'mitoRNA' refers to mitochondrial transcripts. IGR, intergenic region;

miscRNA, miscellaneous RNA. **c**, Compared to extracellular *Salmonella* (0 h), intracellular bacteria at 4 h p.i. repress SPI-1 and induce SPI-2 effector genes. **d**, Invaded (GFP⁺) host cells at 24 h p.i. activate NF- κ B-associated immunity genes. Data in panels **c** and **d** represent fold changes calculated by edgeR over 3 biological replicates.

was consistent with published pathogen²² and host²³ microarray data: expression of invasion-related genes of *Salmonella* pathogenicity island 1 (SPI-1) decreased after bacterial internalization, whereas expression of SPI-2 genes promoting intracellular survival increased (Fig. 1c). In invaded host cells, NF- κ B-associated immunity genes were strongly activated (Fig. 1d). Thus, dual RNA-seq of mixed total RNA reliably profiled both coding and noncoding RNA patterns in intracellular bacteria and their host cells.

Dynamic sRNA expression in intracellular *Salmonella*

To increase coverage of informative transcripts and to make dual RNA-seq more sensitive, we successfully depleted both bacterial and eukaryotic rRNA (Extended Data Fig. 2). We then profiled sRNAs in intracellular *Salmonella* with high resolution, analysing time-course samples taken before and 2, 4, 8, 16 and 24 h after infection of HeLa cells (Extended Data Fig. 3a). Reads per kilobase transcript per million reads (RPKM) distributions revealed the relative abundance of bacterial and human RNA classes (Extended Data Fig. 3b). Altogether, we recorded expression of 145 known and 189 candidate *Salmonella* sRNAs, some of which were already induced greater than tenfold 2 h after invasion (Fig. 2a). Expression changes of well characterized sRNAs provided insight into *Salmonella*'s microenvironment inside the host, as exemplified by RyhB and IsrE which are activated following iron scarcity^{7,24} or MicA/L, RyhB and OmrA/B which all report bacterial surface stress^{25,26}. In addition, decreased SPI-1 and increased SPI-2 expression after bacterial internalization (see Fig. 1c) is reflected in co-transcriptionally regulated sRNAs: InvR and DapZ were repressed^{5,8}, whereas MgrR was activated²⁷.

The most activated sRNA, PinT (renamed from STnc440 (ref. 8) to PhoP-induced sRNA in intracellular *Salmonella*; see later) increased up to 100-fold during the infection, as well as in *Salmonella* grown in SPI-2-inducing medium which mimics the intracellular milieu²⁸ (Fig. 2a, b and Extended Data Fig. 3c). Likewise, PinT was prominently upregulated in dual RNA-seq experiments of 13 other host cell types, including murine bone marrow-derived macrophages, differentiated human THP-1 macrophages, and porcine macrophage-like cells (Fig. 2a, insets; Extended Data Fig. 4a, b).

PinT is an 80 nucleotide-long sRNA^{8,29} from a horizontally acquired *Salmonella*-specific locus that also encodes RtsA, a co-activator protein of invasion genes (Fig. 2c). However, a correlation analysis of all dual RNA-seq time points with global *Salmonella* regulons predicted

control by PhoP/Q (Fig. 2d), the SPI-2-activating two-component system that is essential for intracellular survival^{30,31}. Supporting this prediction, PinT expression was abrogated in a Δ phoP strain and by mutations in the putative PhoP box in the *pinT* promoter (Extended Data Fig. 3d, e). Thus, PinT is a PhoP-dependent *Salmonella* sRNA that is highly activated upon infection of diverse host cells.

PinT sRNA times virulence mRNA expression

PinT exemplifies the current difficulty in understanding sRNA functions in bacterial virulence: an Hfq-bound sRNA of unknown function^{5,8,29}, PinT was selected as a potential *Salmonella* virulence factor in genome-wide random mutagenesis screens (TraDIS) in pigs and cattle¹³, two large-animal models of salmonellosis. Both its sequence conservation (Extended Data Fig. 3f) and its strong induction inside host cells further support a role for this sRNA in bacterial virulence. However, as the *pinT* gene was not selected by TraDIS in mice¹³ and its deletion produces weak macroscopic phenotypes in cultured cells (Extended Data Fig. 3g, h and Extended Data Fig. 4c–e), assays to decipher its molecular function during infection are currently lacking.

Combining several approaches, we discovered a molecular function of PinT as a timer of virulence gene expression, as summarized in Fig. 3a. First, a dual RNA-seq time-course using a Δ pinT strain to infect HeLa cells enabled the prediction that this sRNA represses SPI-2 genes during the early stages after host cell invasion (Fig. 3b and Extended Data Fig. 5a). The PinT-mediated repression of SPI-2 was independently validated for several transcripts for the secretion apparatus and effector proteins and by the rescue of wild-type expression profiles upon sRNA *trans*-complementation (strain *pinT*⁺) (Extended Data Fig. 5b). A dual RNA-seq time-course of *Salmonella*-infected cells from pigs, an organism in which PinT scored as a potential virulence factor¹³, also confirmed the PinT effect on SPI-2 (Fig. 3b and Extended Data Fig. 5a, c). Second, mRNA profiling suggested that PinT acts upstream of the SPI-2 master transcription factor, SsrB, without affecting its own activator PhoP/Q or the invasion gene regulator, HilD (Extended Data Fig. 5b). Third, we successfully recapitulated PinT-mediated repression of SPI-2 genes under defined *in vitro* conditions (Fig. 3b and Extended Data Fig. 5a).

We reasoned that as PinT associates with Hfq²⁹, it regulates mRNAs by base pairing³². To identify early-infection mRNA targets, we pulse-expressed PinT under pre-invasion conditions *in vitro* (Extended Data Fig. 6a and Supplementary Table 1). Two of

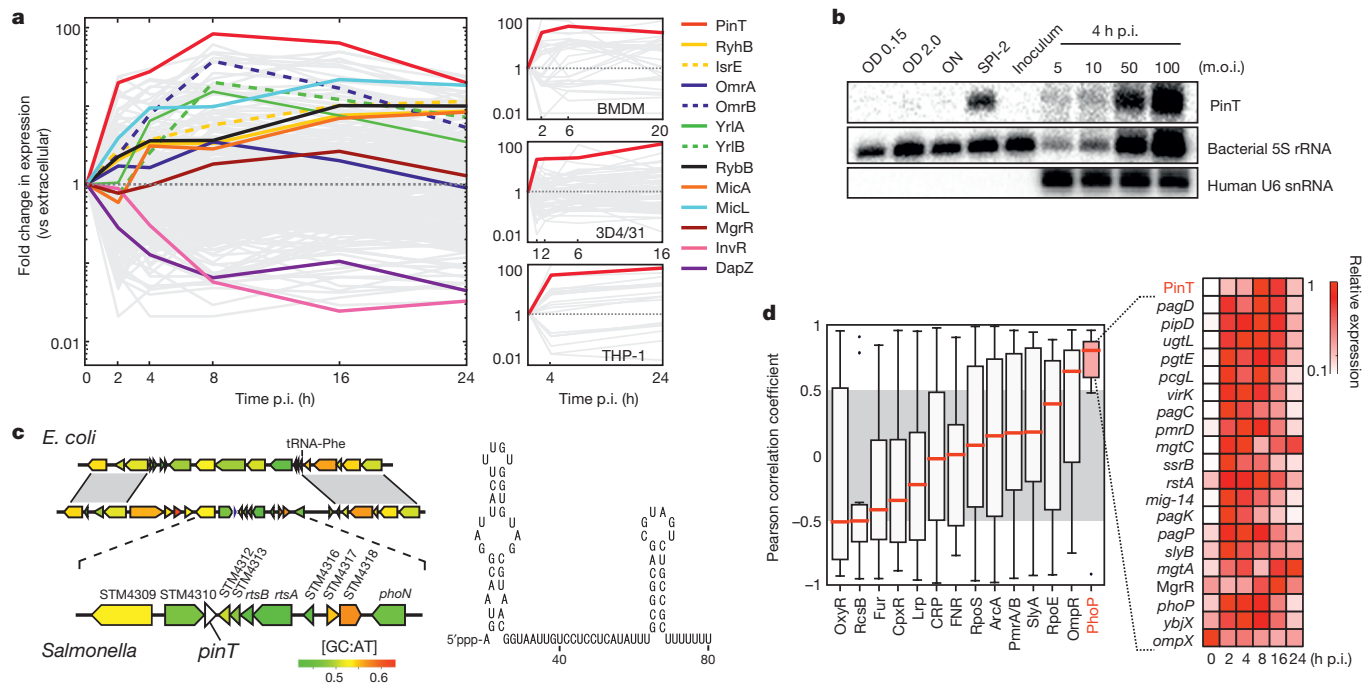


Figure 2 | A PhoP-dependent *Salmonella* sRNA induced inside host cells. **a**, Dual RNA-seq profiles of regulated *Salmonella* sRNAs in HeLa-S3 cells (adjusted P value < 0.05). Insets, PinT sRNA expression (red lines) in different macrophage-like cells. **b**, Northern blot detection of PinT. 5S and U6 RNAs represent loading controls. ON, overnight; 'Inoculum', bacteria in host cell medium before infection. Uncropped gel image in Supplementary Fig. 1. **c**, *pinT* locus and sRNA secondary structure.

the top repressed mRNAs encoded the secreted SPI-1 effectors SopE and SopE2. Conversely, the *sopE/E2* mRNAs were found to be depressed in $\Delta pinT$ *Salmonella* during host cell invasion (Extended Data Fig. 5b). SopE and SopE2 are guanidyl nucleotide exchange factors that stimulate innate immune responses³³ and contribute to the establishment of *Salmonella*'s replicative niche³⁴. They were readily depleted by overexpression of PinT, while other SPI-1 effectors such as SopB or SipC were unaffected (Fig. 3c), arguing for selective regulation. *In silico* modelling predicted that PinT base pairs near the start codon of the *sopE* and *sopE2* mRNAs, which was confirmed by compensatory mutations in *gfp* reporter constructs of these targets, and within the single-stranded 'seed' region of the sRNA (PinT* variant; Extended Data Fig. 6b, c).

To identify relevant PinT targets upon host cell entry, we successfully pulse-expressed PinT in bacteria growing inside host cells and also analysed PinT overexpression in the SPI-2-inducing medium (Extended Data Fig. 6d, e and Supplementary Table 1). The combined data sets enabled the prediction that PinT also directly represses the mRNAs encoding the proteins GrxA (glutathione/glutaredoxin system) and CRP (cyclic AMP receptor protein), by using the same seed sequence as above (Fig. 3c and Extended Data Fig. 6f).

As CRP and GrxA contribute to virulence gene activation in intracellular *Salmonella*^{35–37}, they might mediate PinT signalling to SPI-2. To address this, we monitored regulation of the well-known SsrB-activated SPI-2 gene *ssaG*²⁸ after a switch from SPI-1- to SPI-2-inducing media (Fig. 3d). This *in vitro* assay recapitulated the premature activation of SPI-2 previously seen in intracellular $\Delta pinT$ bacteria (Fig. 3b). Reciprocally, overexpression of PinT impaired SPI-2 gene activation. Importantly, this regulation was lost upon genomic deletion of *crp* but not of *grxA* (Fig. 3d), establishing the metabolic regulator CRP as a mediator of PinT control of SPI-2.

Together these data suggest that PinT shapes the transition from invasion to the intracellular replication state of *Salmonella* by simultaneously acting on two SPI-1 effectors and the SPI-2 virulence genes.

d, PinT as part of the PhoP regulon. Left, 14 analysed global regulons (Supplementary Table 1). Box plots show the interquartile range (IQR), with the median marked (red line). Whiskers indicate the highest/lowest point within $1.5 \times$ IQR of the upper/lower quartile. Right, expression of PhoP target genes during the course of infection. Sequencing data in panels **a** and **d** are derived from 3 biological replicates.

This model was supported with evidence at the protein level of an incomplete clearance of the SPI-1 effector SopE and a faster accumulation of SPI-2 effector SteC in the $\Delta pinT$ bacteria upon shifting from SPI-1 to SPI-2 conditions, as compared to PinT-expressing *Salmonella* (Fig. 3e). Intriguingly, although previous work discovered transcriptional loops that time SPI-2 expression³⁸, our study reveals that PinT provides a post-transcriptional repressor arm of PhoP in a complex feedback loop that helps *Salmonella* transit from SPI-1 to SPI-2 activity upon internalization (Fig. 3a, f). This makes PinT the first sRNA, to our knowledge, known to temporally shape the transition between two major bacterial virulence programs.

PinT impacts the host response

To understand how temporal virulence factor control by PinT may impact the host, we interrogated the dual RNA-seq time-course data for PinT-dependent changes in the transcriptomes of infected HeLa cells. Strikingly, wild-type and $\Delta pinT$ *Salmonella* elicit very different expression patterns amongst the 14,001 mRNAs, 3,982 lncRNAs and 134 miRNAs detected for the host (Fig. 4a and Extended Data Fig. 7a, b). Moreover, while PinT affects bacterial virulence gene expression early post-invasion of HeLa cells (Fig. 3b), its impact on the host transcriptome is apparent throughout the course of the infection (Fig. 4a); even before a small replication phenotype of the $\Delta pinT$ strain has manifested (Extended Data Fig. 3h).

Interspecies correlation analysis of pathogen and host transcriptome changes (see Methods) provided a potential molecular scenario of how PinT affects pathogen–host interplay by linking PinT-regulated SPI-2 gene expression in the bacteria with factors involved in the JAK–STAT pathway in the host (Fig. 4b and Supplementary Table 1). Specifically, we predicted and validated an accelerated activation of a key regulator of JAK–STAT signalling, Suppressor of cytokine signalling 3 (SOCS3), in the absence of PinT (Fig. 4a, c and Extended Data Fig. 7c). SOCS3 inhibits the phosphorylation of the STAT3 (signal transducers and activators of transcription 3) transcription factor to prevent its activation

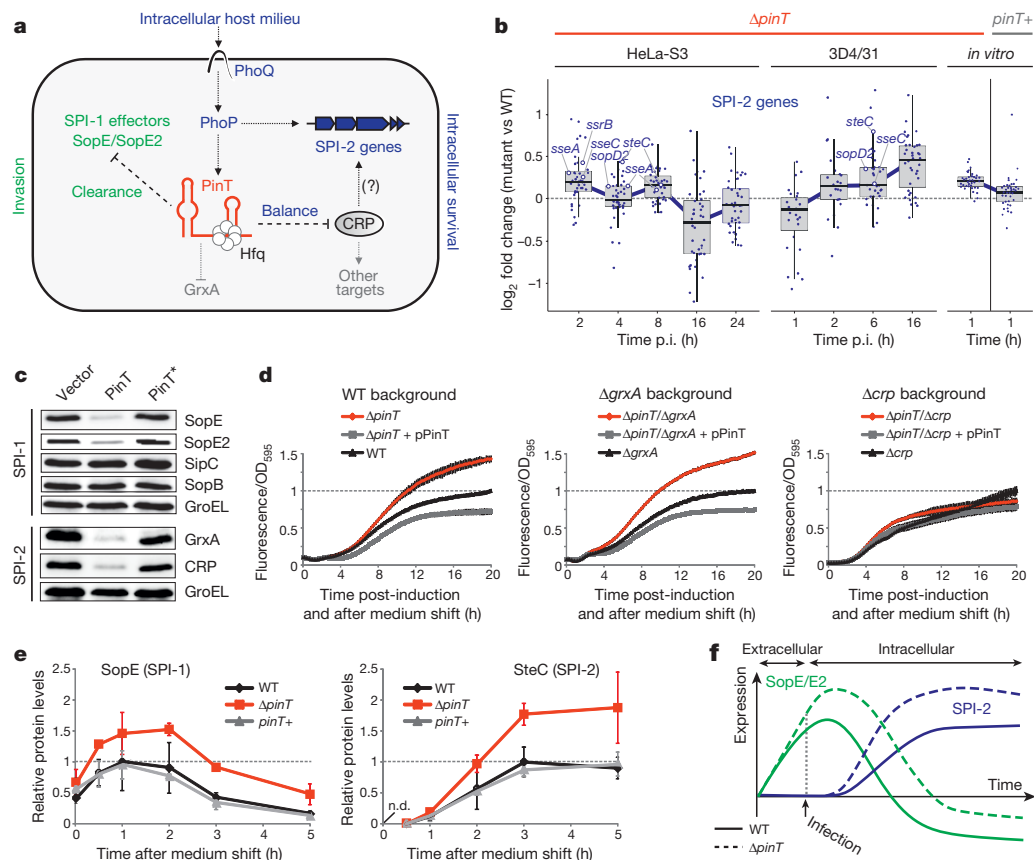


Figure 3 | PinT temporally controls *Salmonella* virulence genes.

a, Model of virulence gene expression regulation by PinT (based on panels **b–e**). Green, SPI-1 branch; blue, SPI-2. **b**, Relative expression of SPI-2 genes between wild-type *Salmonella* and the Δ pinT or pinT+ mutant strains. Box plots represent median expression fold changes of the regulon (whiskers: $1.5 \times$ IQR of the upper/lower quartile). Validated transcripts (Extended Data Fig. 5b, c) are labelled. **c**, Western blot analysis of PinT targets (identified in Extended Data Fig. 6). Overexpression of wild-type (PinT) or point-mutated (PinT*) sRNA under SPI-1 (LB, OD₆₀₀ = 2.0) or SPI-2 conditions (SPI-2 medium, OD₆₀₀ = 0.3). Empty vector and GroEL are negative and loading controls,

respectively. Uncropped gel image in Supplementary Fig. 1. **d**, PinT represses SPI-2 through CRP. *Salmonella* strains with endogenous (black), ectopic (grey) or lacking PinT expression (red) in either wild-type or deletion background were shifted from SPI-1 to SPI-2 conditions (see Methods). A transcriptional reporter fusion of the SPI-2 gene *ssaG* to *gfp* was used as a proxy for SPI-2 induction. **e**, Western blot quantification of *sopE::flag/steC::flag Salmonella* shifted from SPI-1 to SPI-2 conditions. Protein levels were normalized to GroEL. Points and error bars indicate the mean \pm s.d. from 3 biological replicates. n.d., not detected. **f**, Model of the temporal expression of virulence genes in the presence or absence of PinT.

and nuclear import³⁹. Consistent with premature induction of SOCS3 (Fig. 4c), we observed reduced STAT3 phosphorylation in cells infected with Δ pinT *Salmonella* compared to wild-type infection (Fig. 4d).

Dual RNA-seq revealed additional impact of PinT activity on host immune pathways; for example, increased mRNA abundance of the pro-inflammatory chemokine interleukin 8 (*IL8*, also known as *CXCL8*) in HeLa cells infected with Δ pinT bacteria, which was further confirmed on both RNA and protein levels (Extended Data Fig. 7a, d, e). Interestingly, elevated *IL8* mRNA levels spread to bystander cells in Δ pinT-infected cultures (Extended Data Fig. 7d), probably due to paracrine immune signalling⁴⁰. Importantly, both PinT-controlled SPI-1 effectors SopE and SopE2, as well as several SPI-2 effectors influence JAK–STAT signalling or IL-8 secretion^{23,33,41,42}. These host responses are crucial for *Salmonella* to establish its intracellular replication niche²³ and compete with the intestinal microbiota^{43,44}, but also mediate bacterial killing⁴⁵. Therefore, our observations would favour a model in which *Salmonella* uses PinT-mediated temporal control of virulence genes for optimal manipulation of these key host cell pathways to promote its own replication.

Our dual RNA-seq data provides the first, to our knowledge, global map of both polyadenylated and non-polyadenylated host transcripts that respond to a bacterial infection (Extended Data Fig. 7a), providing a temporal view of infection-related lncRNA expression and processing. Intracellular *Salmonella* affect the levels of ~44% of all

detected lncRNAs throughout the genome, almost half of which differ between wild-type and Δ pinT infections (Fig. 4a and Extended Data Fig. 7a, b, f). Remarkably, certain lncRNAs appear to respond very quickly to PinT-dependent alterations (Fig. 4a), suggesting that lncRNAs can provide sensitive markers for pathogen activities in the early infection phase.

Dual RNA-seq further permits the analysis of alternative splicing events, the formation of circular RNAs, and expression changes in small open reading frames; however, PinT generally had little, if any, impact on these processes (data not shown). However, in addition to pathogen and nuclear host transcripts, our protocol readily captures non-polyadenylated organellar transcripts (Fig. 1b), revealing that *Salmonella* strongly induces mitochondrial gene expression in invaded host cells (Extended Data Fig. 7a). Moreover, the Δ pinT strain caused hyperactivation of global mitochondrial RNA expression including the mitochondrial oxidative phosphorylation pathway and altered the subcellular localization of mitochondria (Fig. 4a, e and Extended Data Fig. 8). Although the underlying molecular mechanisms remain to be determined, these findings illustrate how a single sRNA affects host–pathogen interactions at different levels (Fig. 4f). Importantly, the combined data exemplifies how the generic analysis of all transcript classes by dual RNA-seq can reveal changes at the cellular level to provide molecular insight into the roles of genes identified in pathogenesis screens using animal models.

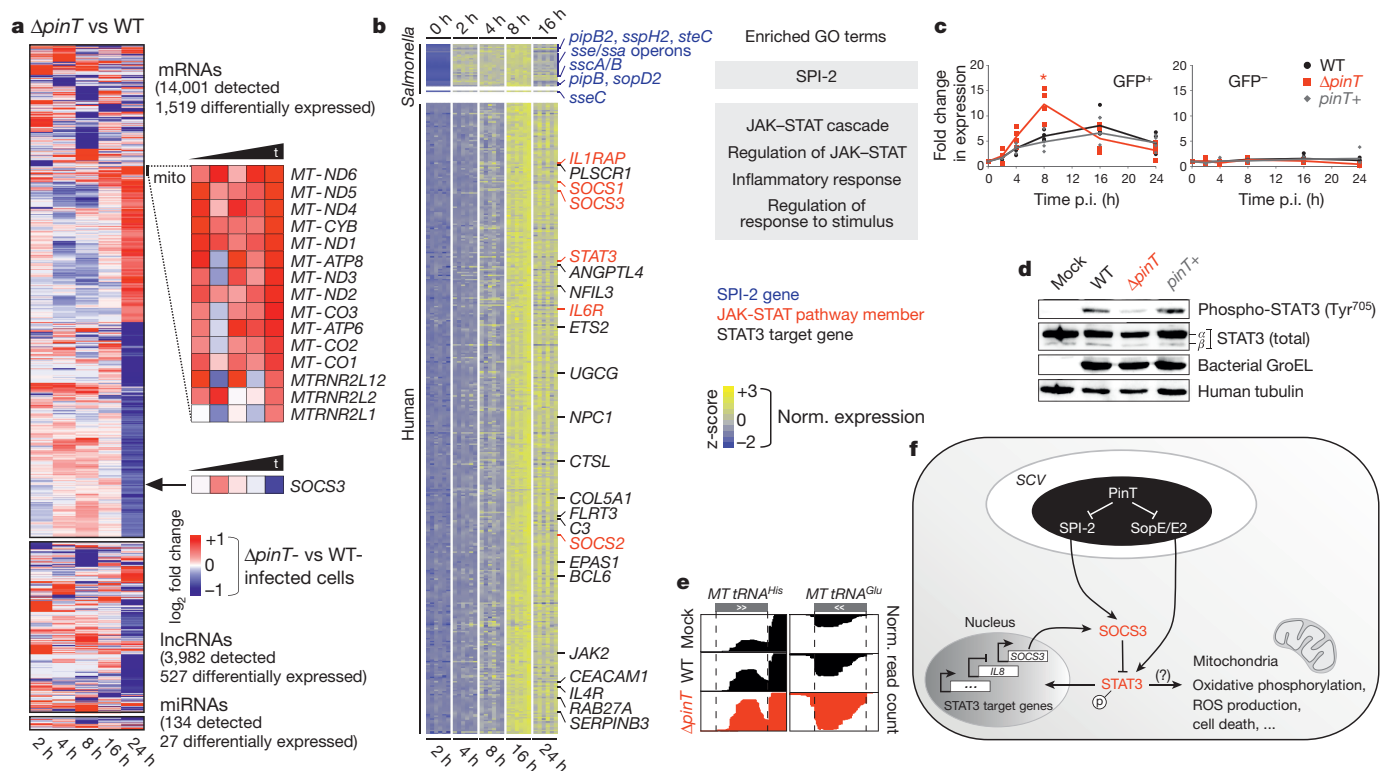


Figure 4 | Effect of PinT activity on the host transcriptome.

a, Differentially expressed HeLa-S3 transcripts (adjusted P value < 0.05 ; 3 biological replicates) between wild-type and $\Delta pinT$ infections. Numbers reflect detected and differentially expressed transcripts, respectively. Full gene lists are in Supplementary Table 1. **b**, Interspecies co-regulation analysis (see Methods) correlates *Salmonella* SPI-2 genes with the human JAK-STAT pathway. Enriched bacterial and human GO terms are displayed. Complete gene sets in Supplementary Table 1. **c**, qRT-PCR measurements of human SOCS3 mRNA in infected (GFP⁺) and bystander (GFP⁻) HeLa-S3 cells after wild-type, $\Delta pinT$ or $pinT^+$ infection. Data represent results from 3 (2h) or 4 (other time points) biological replicates,

Outlook

Using dual RNA-seq, we have comprehensively charted the dynamic RNA expression landscape of both a bacterial pathogen and its eukaryotic host during the course of infection. This approach enabled us to discover PinT as a post-invasion-activated sRNA whose function in host-pathogen interactions manifests itself in pervasive expression changes in all classes of host RNA. Many other sRNA loci are strongly induced upon host cell invasion (Fig. 2a). We selected six whose *in vivo* induction was recapitulated in SPI-2 medium (Extended Data Fig. 9a) for pairwise genomic inactivation. Of these, OmrA/B and RyhB/IsrE each control multiple mRNAs by Hfq-dependent base pairing^{24,26,46}, while YrlA/B program RNA decay in *Salmonella*⁴⁷. Whereas none of the $\Delta omrA/\Delta omrB$, $\Delta ryhB/\Delta isrE$, and $\Delta yrlA/\Delta yrlB$ deletion strains displayed a phenotype in standard invasion/replication assays (Extended Data Fig. 9b, c), initial dual RNA-seq of HeLa cells infected with these double mutants revealed strain-specific changes in *Salmonella* transcripts (Extended Data Fig. 9d, e) and differential pathway activities in infected host cells (Extended Data Fig. 9f, g). While certain host pathways were generically de-regulated by all three mutants (Extended Data Fig. 9g), other pathways were specifically impacted by distinct deletion strains, suggesting that *Salmonella* sRNAs actively contribute to successful host infection by affecting common as well as disparate host pathways.

Virulence factor screens have identified many *Salmonella* protein-coding and sRNA genes whose molecular contributions to successful infection have remained unknown due to a failure to link a phenotype to its underlying mechanism^{12–14,48,49}. Our findings with

asterisk denotes significant expression differences between wild-type- and $\Delta pinT$ -infected cells ($P < 0.05$; one-tailed Mann-Whitney U -test).

d, Western blot of phosphorylation status of STAT3 in unsorted HeLa-S3 cells at 16 h p.i. with the indicated strains, including bacterial (GroEL) and human (tubulin) loading controls. Uncropped gel image in Supplementary Fig. 1. **e**, Representative dual RNA-seq coverage plot for two mitochondrial tRNA genes at 16 h p.i. of HeLa-S3. **f**, Summary of how PinT-dependent regulation of bacterial effectors that affect SOCS3 (ref. 41) or, via Rho-GTPases, STAT3 (ref. 23), may influence host JAK-STAT signalling. STAT3 also affects mitochondria⁵⁰, which potentially interconnects the different PinT-affected host pathways. SCV, *Salmonella*-containing vacuole.

PinT and other intracellularly induced sRNAs illustrate how small perturbations in the infection process, such as dysregulation of a few *Salmonella* mRNAs, can propagate through the entire host system, potentially leading to different disease outcomes in the context of a whole organism. The one-step nature of dual RNA-seq should enable high-throughput studies to unravel such hidden gene functions simultaneously in the pathogen and host, during infection with *Salmonella* and many other pathogens.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2014; accepted 18 December 2015.

Published online 20 January 2016.

- Caldelari, I., Chao, Y., Romby, P. & Vogel, J. RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harb. Perspect. Med.* **3**, <http://dx.doi.org/10.1101/cshperspect.a010298> (2013).
- Gottesman, S. & Storz, G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspect. Biol.* **3**, <http://dx.doi.org/10.1101/cshperspect.a003798> (2011).
- Wagner, E. G. H. & Romby, P. Small RNAs in Bacteria and Archaea: who they are, what they do, and how they do it. *Adv. Genet.* **90**, 133–208 (2015).
- Kroger, C. et al. An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).
- Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M. & Vogel, J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).
- Ortega, A. D., Gonzalo-Asensio, J. & Garcia-del Portillo, F. Dynamics of *Salmonella* small RNA expression in non-growing bacteria located inside eukaryotic cells. *RNA Biol.* **9**, 469–488 (2012).

7. Padalon-Brauch, G. *et al.* Small RNAs encoded within genetic islands of *Salmonella* typhimurium show host-induced expression and role in virulence. *Nucleic Acids Res.* **36**, 1913–1927 (2008).
8. Pfeiffer, V. *et al.* A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella* core genome. *Mol. Microbiol.* **66**, 1174–1191 (2007).
9. Gong, H. *et al.* A *Salmonella* small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog.* **7**, e1002120 (2011).
10. Papenfert, K., Podkaminski, D., Hinton, J. C. & Vogel, J. The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G–U wobble pair. *Proc. Natl Acad. Sci. USA* **109**, E757–E764 (2012).
11. Sittka, A., Pfeiffer, V., Tedin, K. & Vogel, J. The RNA chaperone Hfq is essential for the virulence of *Salmonella* typhimurium. *Mol. Microbiol.* **63**, 193–217 (2007).
12. Santiviago, C. A. *et al.* Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog.* **5**, e1000477 (2009).
13. Chaudhuri, R. R. *et al.* Comprehensive assignment of roles for *Salmonella* Typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genetics* **9**, e1003456 (2013).
14. Barquist, L. & Vogel, J. Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu. Rev. Genet.* **49**, 367–394 (2015).
15. Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nature Rev. Microbiol.* **10**, 618–630 (2012).
16. Sorek, R. & Cossart, P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature Rev. Genet.* **11**, 9–16, (2010).
17. Humphrys, M. S. *et al.* Simultaneous transcriptional profiling of bacteria and their host cells. *PLoS ONE* **8**, e80597 (2013).
18. Malik-Kale, P. *et al.* *Salmonella*—at home in the host cell. *Front. Microbiol.* **2**, 125 (2011).
19. Wyllie, D. H. *et al.* Evidence for an accessory protein function for Toll-like receptor 1 in anti-bacterial responses. *J. Immunol.* **165**, 7125–7132 (2000).
20. Bumann, D. Examination of *Salmonella* gene expression in an infected mammalian host using the green fluorescent protein and two-colour flow cytometry. *Mol. Microbiol.* **43**, 1269–1283 (2002).
21. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**, 250–255 (2010).
22. Hautefort, I. *et al.* During infection of epithelial cells *Salmonella enterica* serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell Microbiol.* **10**, 958–984 (2008).
23. Hannemann, S., Gao, B. & Galan, J. E. *Salmonella* modulation of host cell gene expression promotes its intracellular growth. *PLoS Pathog.* **9**, e1003668 (2013).
24. Masse, E., Vanderpool, C. K. & Gottesman, S. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J. Bacteriol.* **187**, 6962–6971 (2005).
25. Guo, M. S. *et al.* MicL, a new σ -E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* **28**, 1620–1634 (2014).
26. Guillier, M. & Gottesman, S. The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Res.* **36**, 6781–6794 (2008).
27. Moon, K., Six, D. A., Lee, H. J., Raetz, C. R. & Gottesman, S. Complex transcriptional and post-transcriptional regulation of an enzyme for lipopolysaccharide modification. *Mol. Microbiol.* **89**, 52–64 (2013).
28. Löber, S., Jackel, D., Kaiser, N. & Hensel, M. Regulation of *Salmonella* pathogenicity island 2 genes by independent environmental signals. *Int. J. Med. Microbiol.* **296**, 435–447 (2006).
29. Sittka, A. *et al.* Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* **4**, e1000163 (2008).
30. Prost, L. R., Daley, M. E., Bader, M. W., Klevit, R. E. & Miller, S. I. The PhoQ histidine kinases of *Salmonella* and *Pseudomonas* spp. are structurally and functionally different: evidence that pH and antimicrobial peptide sensing contribute to mammalian pathogenesis. *Mol. Microbiol.* **69**, 503–519 (2008).
31. Groisman, E. A. The pleiotropic two-component regulatory system PhoP–PhoQ. *J. Bacteriol.* **183**, 1835–1842 (2001).
32. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nature Rev. Microbiol.* **9**, 578–589 (2011).
33. Bruno, V. M. *et al.* *Salmonella* Typhimurium type III secretion effectors stimulate innate immune responses in cultured epithelial cells. *PLoS Pathog.* **5**, e1000538 (2009).
34. Vonaesch, P. *et al.* The *Salmonella* Typhimurium effector protein SopE transiently localizes to the early SCV and contributes to intracellular replication. *Cell. Microbiol.* **16**, 1723–1735 (2014).
35. Bjur, E., Eriksson-Ygberg, S., Aslund, F. & Rhen, M. Thioredoxin 1 promotes intracellular replication and virulence of *Salmonella enterica* serovar Typhimurium. *Infect. Immun.* **74**, 5140–5151 (2006).
36. Chen, Z. W. *et al.* Mutations in the *Salmonella enterica* serovar Choleraesuis cAMP-receptor protein gene lead to functional defects in the SPI-1 type III secretion system. *Vet. Res.* **41**, 05 (2010).
37. Yoon, H. J., McDermott, J. E., Porwollik, S., McClelland, M. & Heffron, F. Coordinated regulation of virulence during systemic infection of *Salmonella enterica* serovar Typhimurium. *PLoS Pathogens* **5**, e1000306 (2009).
38. Choi, J. & Groisman, E. A. The lipopolysaccharide modification regulator PmrA limits *Salmonella* virulence by repressing the type three-secretion system Spi/Ssa. *Proc. Natl Acad. Sci. USA* **110**, 9499–9504 (2013).
39. Li, Y., de Haar, C., Peppelenbosch, M. P. & van der Woude, C. J. SOCS3 in immune regulation of inflammatory bowel disease and inflammatory bowel disease-related cancer. *Cytokine Growth Factor Rev.* **23**, 127–138 (2012).
40. Browning, D. D., Diehl, W. C., Hsu, M. H., Schraufstatter, I. U. & Ye, R. D. Autocrine regulation of interleukin-18 production in human monocytes. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **279**, L1129–L1136 (2000).
41. Uchiya, K. & Nikai, T. *Salmonella* pathogenicity island 2-dependent expression of suppressor of cytokine signaling 3 in macrophages. *Infect. Immun.* **73**, 5587–5594 (2005).
42. Bhavsar, A. P. *et al.* The *Salmonella* type III effector SspH2 specifically exploits the NLR co-chaperone activity of SGT1 to subvert immunity. *PLoS Pathog.* **9**, e1003518 (2013).
43. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* **467**, 426–429 (2010).
44. Stecher, B. *et al.* *Salmonella enterica* serovar Typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol.* **5**, e244 (2007).
45. Gewirtz, A. T., Siber, A. M., Madara, J. L. & McCormick, B. A. Orchestration of neutrophil movement by intestinal epithelial cells in response to *Salmonella* typhimurium can be uncoupled from bacterial internalization. *Infect. Immun.* **67**, 608–617 (1999).
46. Holmqvist, E. *et al.* Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J.* **29**, 1840–1850 (2010).
47. Chen, X. *et al.* An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. *Cell* **153**, 166–177 (2013).
48. Mazurkiewicz, P., Tang, C. M., Boone, C. & Holden, D. W. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nature Rev. Genet.* **7**, 929–939 (2006).
49. Becker, D. *et al.* Robust *Salmonella* metabolism limits possibilities for new antimicrobials. *Nature* **440**, 303–307 (2006).
50. Myers, M. G. Moonlighting in mitochondria. *Science* **323**, 723–724 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Gorski for help with the manuscript, C. Sharma and A. Eulalio for comments on the paper and L. Pfeuffer, B. Plaschke and V. McParland for technical assistance. We further thank S. Hoffmann for discussions about bioinformatic analyses, T. Rudel, K. Tedin, T. Meyer and C. Sharma for eukaryotic cell lines, V. Kozjak-Pavlovic for MitoTracker orange dye, T. Strowig for Ribo-Zero reagents, and W.-D. Hardt, M. Kolbe, H. Aiba and A. Eulalio for providing primary antibodies. A.J.W. was the recipient of an Elite Advancement Ph.D. stipend from the Universität Bayern e.V., Germany. L.B. is supported by a research fellowship from the Alexander von Humboldt Stiftung/Foundation. The Vogel laboratory received funding from the Bavarian BioSysNet program and BMBF (Bundesministerium für Bildung und Forschung) grant “Next-generation transcriptomics of bacterial infections”. This work was further supported by a donation from Baldwin Knauf to the Research Center for Infectious Diseases (ZINF). J.V. and P.F.S. acknowledge support by a joint BMBF grant eBio:RNAsys. This work used the European Grid Infrastructure (EGI) and funding by the EGI-Engage project (Horizon 2020) under grant number 654142.

Author Contributions A.J.W., Y.C., L.N.S. and J.V. designed the research; A.J.W., Y.C. and L.N.S. performed the bench laboratory work; K.U.F., F.A., L.B., L.M. and P.F.S. conducted computational analyses; R.R. carried out part of the sequencing; A.J.W. and J.V. wrote the manuscript, which all co-authors commented on. J.V. supervised the project.

Author Information All RNA-seq data has been deposited in NCBI's Gene Expression Omnibus under accession number GSE60144. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.V. (joerg.vogel@uni-wuerzburg.de).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Salmonella strains and mammalian cell lines. *Salmonella enterica* serovar Typhimurium strain SL1344 constitutively expressing GFP from a chromosomal locus (strain JVS-3858) was previously described⁵¹ and is referred to as wild type throughout this study. The complete list of bacterial strains used in this study is provided in Supplementary Table 1. Routinely, bacteria were grown in Lennox broth (LB) medium at 37°C with shaking at 220 r.p.m. When appropriate, 100 µg ml⁻¹ ampicillin (Amp), 50 µg ml⁻¹ kanamycin (Kan), or 20 µg ml⁻¹ chloramphenicol (Cm) (final concentrations) were added to the liquid medium or agar plates. Chromosomal mutagenesis of *Salmonella* SL1344 was performed as previously described⁵². To construct a non-polar *pinT* mutant strain (YCS-034, GFP⁻; or JVS-10038, GFP⁺), the first ~60 nt of the gene were removed and replaced by a resistance cassette, while keeping the Rho-independent terminator intact. Then, the resistance cassette was eliminated using the FLP helper plasmid pCP20 at 42°C⁵². All mutations were transduced into the wild-type background using P22 phage⁵³. For plasmid transformation the respective *Salmonella* strains were electroporated with ~10 ng of DNA.

The following cell lines were used in this study: human cervix carcinoma cells (HeLa-S3; ATCC CCL-2.2), human epithelial colorectal adenocarcinoma cells (CaCo-2; ATCC HTB-37), human epithelial colorectal adenocarcinoma cells (HT29; DSMZ No. ACC-299), human stomach adenocarcinoma cells (AGS; ATCC CRL-1739), human epithelial colon metastatic cells (LoVo; ATCC CCL-229), human embryonic kidney 293 cells (HEK293; ATCC CRL-1573), human monocytic cells (THP-1; ATCC TIB-202), murine fibroblast cells (L929; ATCC CCL-1), murine embryonic fibroblast cells (MEF; ATCC SCRC-1040), mouse leukaemic monocyte/macrophage cells (RAW264.7; ATCC TIB-71), porcine intestinal epithelial cells (IPEC-J2)⁵⁴, porcine macrophage-like cells (3D4/31)⁵⁵.

HeLa-S3, CaCo-2, THP-1, HEK293, RAW264.7 and MEF cells were obtained from the group of Thomas Rudel (Biocentre, Würzburg). AGS cells were provided by Cynthia Sharma (Research Centre for Infectious Diseases, Würzburg). L929 cells were obtained from Thomas Meyer (Max Planck Institute for Infection Biology, Berlin). HT29, LoVo, IPEC-J2 and 3D4/31 cells were provided by Karsten Tedin (Centre for Infection Medicine, Berlin). Cell lines have not been authenticated in our laboratory, but were routinely tested for mycoplasma contamination (MycoAlert Mycoplasma Detection Kit, Lonza).

HeLa-S3 cells were cultured according to the guidelines provided by the ENCODE consortium (http://genome.ucsc.edu/encode/protocols/cell/human/Stam_15_protocols.pdf). Briefly, cells were grown in DMEM (Gibco) supplemented with 10% fetal calf serum (FCS; Biochrom), 2 mM L-glutamine (Gibco) and 1 mM sodium pyruvate (Gibco) in T-75 flasks (Corning) in a 5% CO₂, humidified atmosphere, at 37°C. Further cell lines used in this study (THP-1, CaCo-2, AGS, HT29, LoVo, HEK293, MEF, L929, RAW264.7, IPEC-J2 and 3D4/31) were cultured in RPMI (Gibco) supplemented with 10% FCS, 2 mM L-glutamine, 1 mM sodium pyruvate and 0.5% β-mercaptoethanol (Gibco) in a 5% CO₂, humidified atmosphere, at 37°C. To differentiate THP-1 monocytes, seeded cells (1 × 10⁶ cells per well; six-well format) were treated with 50 ng ml⁻¹ (final concentration) of phorbol 12-myristate 13-acetate (PMA) (Sigma) for 72 h (after 48 h fresh PMA at the same concentration was added to the culture).

For the differentiation of murine bone marrow derived macrophages (BMDMs), the marrow of femur and tibia was isolated from 8–12-week-old female C57BL/6 wild-type mice and stored in RPMI supplemented with 10% FCS. The cell suspension was centrifuged for 5 min at 250g and the leukocyte pellet was resuspended in differentiation medium consisting of X-vivo-15 medium (Lonza) supplemented with 10% FCS and 10% L929-conditioned DMEM medium (same composition as above). Cells were cultured at 3 × 10⁶ cells per 10 ml in a T-75 flask. At day 3, another 3 ml of differentiation medium were added and cells were further cultured until day 5. Successful macrophage differentiation was validated by microscopy before the cells were detached using a rubber scraper (Sarstedt) and seeded into six-well plates at 10⁵ cells per well in fresh differentiation medium. Infection was carried out on day 7 as described below.

Salmonella infection assay. *In vitro* infection of HeLa-S3 cells was carried out following a previously published protocol⁵⁶ with slight modifications. Two days before infection 2 × 10⁵ HeLa-S3 cells were seeded in 2 ml complete DMEM (six-well format). Overnight cultures of *Salmonella* were diluted 1:100 in fresh LB medium and grown aerobically to an OD₆₀₀ of 2.0. Bacterial cells were harvested by centrifugation (2 min at 12,000 r.p.m., room temperature) and resuspended in DMEM. Infection of HeLa-S3 cells was carried out by adding the bacterial suspension directly to each well. If not mentioned otherwise, infections were performed at a multiplicity of infection (m.o.i.) of 5. Immediately after addition of bacteria, the plates were centrifuged for 10 min at 250g at room temperature followed by

30 min incubation in 5% CO₂, humidified atmosphere, at 37°C. Medium was then replaced for gentamicin-containing DMEM (final concentration: 50 µg ml⁻¹) to kill extracellular bacteria. After a further 30 min incubation step, medium was again replaced by fresh DMEM containing 10 µg ml⁻¹ of gentamicin, and incubated for the remainder of the experiment. Time point 0 was defined as the time when gentamicin was first added to the cells.

Further cell types were infected as described for HeLa-S3 cells except that infection was carried out in RPMI medium and that infection was with an m.o.i. of 10 (THP-1, CaCo-2, HT29, AGS, HEK293, MEF, L929 and RAW264.7) or 20 (IPEC-J2, 3D4/31), respectively. Infection of BMDMs was carried out with an m.o.i. of 10 and using X-vivo-15 medium (10% fetal calf serum, 10% L929-conditioned medium).

Confocal laser scanning microscopy. Infection was carried out as described above, except that HeLa-S3 cells had been seeded onto coverslips (24-well format). At the respective timepoint, coverslips with infected HeLa-S3 were washed twice with PBS (Gibco) and fixed in 4% paraformaldehyde (PFA) for 15 min in a wet chamber. After two additional PBS washing steps, cells were stained with Hoechst 33342 (Invitrogen; diluted 1:5,000 in PBS) for 15 min in a wet chamber and again washed twice with PBS. After coverslips had been air-dried, they were embedded in Vectashield Mounting Medium (Biozol) and analysed using the Leica SP5 confocal microscope (Leica) and the LAS AF Lite software (Leica).

To stain human mitochondria, MitoTracker Orange CMTMRos (Life Technologies; kindly provided by V. Kozjak-Pavlovic, Biocentre, Würzburg) was used. The dye was added in the dark to a final concentration of 200 nM directly into the medium of the infected cells in the 37°C incubator, 30 min before their harvest. After the 30 min incubation with the dye, the plates were covered with aluminium foil to prevent bleaching during the following steps. The supernatant was aspirated and the cells were washed with PBS and fixed with 4% PFA at 4°C overnight. Hoechst staining and sample preparation was performed as described above.

Flow cytometry and fluorescence-activated cell sorting (FACS). For flow cytometry-based analyses, infected cultures were washed twice with PBS, detached from the bottom of the plate by trypsinization and resuspended in complete DMEM. Upon pelleting the cells (5 min at 250g, room temperature), they were resuspended in PBS and analysed by flow cytometry using a FACSCalibur instrument (BD Biosciences) and the Cyflogic (CyFlo Ltd; version 1.2.1) or Flowing (Cell Imaging Core, Turku Centre for Biotechnology, Finland; version 2.5.0) software, respectively. Selection of intact HeLa-S3 cells was achieved by gating based on cell diameter (forward-scatter) and granularity (side-scatter) (linear scale). Of those, infected (GFP-positive) and non-infected (GFP-negative) sub-fractions were defined based on GFP signal intensity (FITC channel) versus auto-fluorescence (PE channel) (logarithmic scale).

For cell sorting, RNA later-fixed cells (see below) were first passed through MACS Pre-Separation Filters (30 µm exclusion size; Miltenyi Biotec) and then analysed and sorted using the FACS Aria III device (BD Biosciences) at 4°C (cooling both the input tube holder and the collection tube rack) and at a medium flow rate using the same gating strategy as described above, except that the gates for GFP-positive and GFP-negative fractions were conservative in order to prevent cross-contamination (as exemplified in Extended Data Fig. 1d). Typically ~2 × 10⁵ cells of each fraction were collected for RNA isolation.

Staining of apoptotic cells and cytotoxicity assay. To detect apoptotic cells, HeLa-S3 cells were washed twice with PBS and resuspended in 1 × binding buffer (BD Pharmingen) to a concentration of 10⁶ cells per ml. 100 µl of this cell suspension were mixed with 5 µl of APC-labelled annexin V (BD Pharmingen) and 1 µl of 500 mg ml⁻¹ propidium iodide (PI; lyophilized stock from Sigma). Upon incubation for 15 min at room temperature, (light-protected) cells were subjected to flow cytometry using the MACSQuant Analyzer (Miltenyi Biotec). Upon gating of the fraction of intact cells based on cell diameter (forward-scatter) and granularity (side-scatter), the annexin-positive/PI-negative sub-population was determined by comparison against the appropriate single-stained controls in the APC vs PerCP channels, and quantified.

Necrosis was evaluated by quantifying released lactate dehydrogenase (LDH) via the Cytotox96 assay (Promega) according to the manufacturer's instructions. The absorbance at 490 nm was measured using a Multiskan Ascent instrument (Thermo Fisher). In order to convert the measured absorbance values into the relative proportion of dead cells, the maximal absorbance was determined by using 1 × lysis solution (Promega) following the manufacturer's instructions and referred to as 100% cytotoxicity. For both apoptosis and cytotoxicity measurements each biological replicate comprised three technical replicates.

Quantification of intracellular replication (flow cytometry, c.f.u. assay). To quantify bacterial intracellular replication (Extended Data Fig. 1b), infected host cells were analysed by flow cytometry as described above, except that the increase in GFP intensity (geometric mean) was measured in the GFP-positive

sub-population over time and normalized to that of the non-infected population in the same sample (example in Extended Data Fig. 1c).

Alternatively, infected HeLa-S3 cultures were solubilized with PBS containing 0.1% Triton X-100 (Gibco) at the respective time points. Cell lysates were serially diluted in PBS, plated onto LB plates and incubated at 37°C overnight. The number of colony forming units (c.f.u.) recovered was compared to that obtained from the bacterial input solution used for infection. In all cases, each biological replicate comprised three technical replicates.

Evaluation of different fixation techniques. Infected cells were washed twice with PBS, trypsinized and pelleted. For ethanol fixations, cell pellets were re-dissolved in 0.1 volume of ice-cold PBS and then 0.9 volume of ice-cold ethanol (either 70% or 100%; as indicated) were added in single droplets during shaking (400 r.p.m., 4°C) to avoid cell clumping. Fixation using stop solution (95% EtOH/5% water-saturated phenol)⁵⁷ was performed by resuspending the cell pellet in PBS before the addition of 0.2 volume of stop solution and mixing. When PFA was used, the pellet was resuspended in the respective PFA concentration (0.5% or 4% PFA, pH 7.4, with or without 4% sucrose) and shaken for 15 min at 400 r.p.m., room temperature. PFA-induced crosslinks were reverted by an additional heating step for 15 min at 70°C (refs 58, 59). For fixation with RNeasy (Qiagen), cell pellets were directly resuspended in RNeasy (1 ml per 5×10^6 cells). For systematic evaluation of different fixation protocols (Extended Data Fig. 1e–g), fixed cells had not been sorted but were either directly analysed upon fixation (30 min) or stored at –20°C (ethanol-based fixatives) or 4°C (others), respectively, overnight. To prepare RNeasy-fixed samples for sorting, tubes containing $\sim 5 \times 10^6$ fixed cells were filled up with 10 ml of ice-cold PBS, centrifuged (5 min, 500g, 4°C) and cell pellets resuspended in 2 ml of cold PBS. This cell suspension was filtered and sorted (as described above).

RNA extraction, DNase treatment, evaluation of RNA quality and qRT-PCR. In the dual RNA-seq experiments, as a reference for gene expression changes in host cells upon infection, a non-infected yet mock-treated control was included. The bacterial reference samples were derived from *Salmonella* grown in LB to an OD₆₀₀ of 2.0, which either were then shifted to DMEM for 15 min, pelleted and fixed in RNeasy (see above) or were fixed directly (that is, without a medium exchange step) as indicated. Fixed *Salmonella* cells were pelleted and lysed using the lysis/binding buffer of the mirVana kit (Ambion). In order to maintain the approximate ratio of bacterial to host transcripts during RNA isolation, *Salmonella* lysates were mixed with host cell lysate in a way that the calculated proportion of individual *Salmonella* cells per infected host cell at the latest time point (see Extended Data Fig. 1h) was matched. The resulting mixture was then processed collectively. RNA was extracted from cells using the mirVana kit (Ambion) following the manufacturer's instructions for total RNA isolation. To remove contaminating genomic DNA, samples were treated with 0.25 U of DNase I (Fermentas) per 1 µg of RNA for 45 min at 37°C. If applicable, RNA quality was checked on the Agilent 2100 Bioanalyzer (Agilent Technologies).

For qRT-PCR experiments total RNA was isolated using the TRIzol LS reagent (Invitrogen) according to the manufacturer's recommendations and treated with DNase I (Fermentas) as described above. qRT-PCR was performed with the Power SYBR Green RNA-to-CT 1-Step kit (Applied Biosystems) according to the manufacturer's instructions. Fold changes were determined using the $2^{-\Delta\Delta C_t}$ method⁶⁰. Primer sequences are given in Supplementary Table 1 and their specificity had been confirmed using Primer-BLAST (NCBI). For the estimation of *Salmonella* RNA within infection samples (Extended Data Fig. 1h), a dilution series of separately isolated *Salmonella* and HeLa-S3 total RNA was set up and in each case the ratio of *rfaH*/*ACTB* mRNAs was determined. The same was done for biological samples from infected cells as well as for the *Salmonella* reference controls. From the resulting trend-line equation the approximate proportion of the *Salmonella* transcriptome within mixed prokaryotic and eukaryotic total RNA samples could be deduced.

rRNA depletion (Ribo-Zero treatment). Where indicated (Supplementary Table 1), *Salmonella* and eukaryotic host rRNA were removed using the Ribo-Zero Magnetic Gold Kit (Epidemiology) purchased from Epicentre/Illumina. Following the manufacturer's instructions, ~ 500 ng of total, DNase-I-treated RNA from infection samples was used as an input to the ribosomal transcript removal procedure. rRNA-depleted RNA was precipitated in ethanol for 3 h at –20°C.

cDNA library generation and (dual) RNA-seq. cDNA libraries for Illumina sequencing were generated by Vertis Biotechnologie AG, Freising-Weihenstephan, Germany. For dual RNA-seq of total RNA, at least 100 ng RNA were used for cDNA library preparation. DNase-I-treated total RNA samples were first sheared via ultra-sound sonication (4 pulses of 30 s at 4°C each) to generate ~ 200 –400 bp (average) fragmentation products. Fragments <20 nt were removed using the Agencourt RNAClean XP kit (Beckman Coulter Genomics). As an internal quality control for the pilot experiment (shown in Fig. 1), spike-in RNA (5'-AAAUCCGUUCGUACGGGCC-3'; 5'-monophosphorylated and

gel-purified) was added to a final concentration of 0.5%. The samples were poly(A)-tailed using poly(A) polymerase and the 5' triphosphate (or eukaryotic 5' cap) structures were removed using tobacco acid pyrophosphatase (TAP). Afterwards, an RNA adaptor was ligated to the 5' monophosphate of the RNA fragments. First-strand cDNA synthesis was performed using an oligo(dT)-adaptor primer and the M-MLV reverse transcriptase (NEB). The resulting cDNA was PCR-amplified to about 20–30 ng µl⁻¹ using a high fidelity DNA polymerase (barcode sequences for multiplexing were part of the 3' primers). The cDNA library was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analysed by capillary electrophoresis (Shimadzu MultiNA microchip electrophoresis system).

cDNA libraries for dual RNA-seq on rRNA-depleted samples were constructed as described above, except for the following modifications. Upon RNA fragmentation, dephosphorylation with Antarctic Phosphatase (AP, NEB) and re-phosphorylation with T4 Polynucleotide Kinase (PNK, NEB) were performed. Oligonucleotide adapters were ligated to both the 5' and 3' ends of the RNA samples. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adaptor as primer.

cDNA libraries from *Salmonella*-only samples were generated by fragmenting 5 µg of total RNA using ultrasound and RNAs <20 nt were removed using the Agencourt RNAClean XP kit (Beckman Coulter Genomics) as above. The RNA samples were poly(A)-tailed and 5'ppp structures were removed as before. RNA adapters were ligated to the 5' monophosphate of the RNA and first-strand cDNA synthesis was performed using an oligo(dT)-adaptor primer and the M-MLV reverse transcriptase. The resulting cDNAs were PCR-amplified, purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analysed by capillary electrophoresis (Shimadzu MultiNA microchip).

Generally, for sequencing cDNA samples were pooled in approximately equimolar amounts. The cDNA pool was size-fractionated in the size range of 150–600 bp using a differential clean-up with the Agencourt AMPure kit. For the dual RNA-seq pilot experiment (Fig. 1), single-end sequencing (100 cycles) was performed on an Illumina HiSeq 2000 machine at the Max Planck Genome Centre Cologne, Cologne, Germany. For dual RNA-seq on rRNA-free samples as well as for conventional RNA-seq of *Salmonella*-only samples, single-end sequencing (75 cycles) was performed on a NextSeq500 platform at Vertis Biotechnologie AG, Freising-Weihenstephan, Germany.

All RNA-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE60144. For the accession numbers of individual experiments, see Supplementary Table 1.

Northern blotting (of bacterial and human transcripts). Total RNA prepared with TRIzol LS reagent (Invitrogen) was separated in 6% (vol/vol) polyacrylamide-8.3 M urea gels and blotted as described¹¹. We loaded per lane either 5–10 µg of RNA from pure bacterial samples (Extended Data Figs 3d and 9a), 2 µg total RNA from sorted cell samples (Extended Data Fig. 8b), or 50 µg total RNA from unsorted infection samples (Fig. 2b). Hybond XL membranes (Amersham) were hybridized at 42°C with gene-specific [³²P] end-labelled DNA oligonucleotides (see Supplementary Table 1 for sequences) in Hybri-Quick buffer (Carl Roth AG). **Mutational analysis of the *pinT* promoter region.** The *pinT* promoter region was amplified by PCR using primers JVO-7036/-7037 and inserted via the *AatII* and *NheI* sites in the backbone of plasmid pAS093, resulting in plasmid pYC65. To identify the PhoP binding sites in a minimal fragment, the *pinT* promoter region was truncated by amplifying pYC65 using Phusion polymerase (NEB) with JVO-9393/-7387. The critical residues in the PhoP binding motif (T₋₂₇T₋₂₈) were mutated to adenines by site-directed mutagenesis with JVO-12461/-12462 and Phusion polymerase (NEB).

sRNA pulse-expression (in vitro and in vivo). For pulse-expression of PinT in *in vitro* grown *Salmonella*, we used arabinose-induced overexpression of PinT from a pBAD plasmid previously described^{10,51,61} with minor modifications. Briefly, wild-type *Salmonella* that carried either a pKP8-35 (pBAD control), pYC5-34 (pBAD-PinT) or pYC60 (pBAD-PinT*) plasmid were grown overnight in LB and, the next day, the cultures were 1:100 diluted and further grown in LB to an OD₆₀₀ of 2.0. L-arabinose (Sigma) was added to a final concentration of 0.2%; 5 min later RNA was extracted using TRIzol LS reagent (Invitrogen) and analysed by RNA-seq (~ 3 –5 million reads/library). For the same experiment under SPI-2-inducing conditions, overnight cultures of the three strains were washed 2× with PBS and 1× with SPI-2 medium²⁸, diluted 1:50 in SPI-2 medium and grown to an OD₆₀₀ of 0.3 before PinT expression was induced as above.

For the pulse-expression of PinT inside host cells (Extended Data Fig. 6d, e), HeLa-S3 cells were infected with the same three strains as above and 4 h after infection, 0.2% L-arabinose was supplemented directly into the DMEM medium. Activation of inducible sRNA expression in intracellular bacteria was confirmed by qRT-PCR over a time-course of 20 min (Extended Data Fig. 6d), demonstrating full induction levels to be reached already at 5 min. Thus, for Extended Data Fig. 6e

the host cells were lysed at 5 min after induction with ice-cold 0.1% Triton X-100/PBS and further incubated for 30 min on ice with pipetting up and down from time to time to improve host cell lysis efficiency. Then the intact bacterial cells were pelleted by centrifugation for 2 min at 16,100g (4 °C) and resuspended in RNeasy lysis buffer (Qiagen). The fixed bacterial cells were further enriched against the host background via cell sorting (FACSARIA III, BD Biosciences) and selective gating for the fraction of GFP⁺ bacterial cells released from their hosts. From those, total RNA was isolated and analysed by RNA-seq as above except that sequencing was to a depth of ~20 million reads per library as necessitated by remaining host-derived RNA fragments.

Western blotting (of bacterial and human proteins). Immunoblotting of *Salmonella* proteins was done as previously described⁶². Briefly, samples from *Salmonella* *in vitro* cultures were taken corresponding to 0.4 OD₆₀₀, centrifuged for 4 min at 16,100g at 4 °C, and pellets resuspended in sample loading buffer to a final concentration of 0.01 OD per µl. After denaturation for 5 min at 95 °C, 0.05-OD equivalents of the sample were separated via SDS-PAGE. Gel-fractionated proteins were blotted for 90 min (0.2 mA per cm²; 4 °C) in a semi-dry blotter (Peglab) onto a PVDF membrane (Perkin Elmer) in transfer buffer (25 mM Tris base, 190 mM glycine, 20% methanol). Blocking was for 1 h at room temperature in 10% dry milk/TBST20. Appropriate primary antibodies (see Supplementary Table 1) were hybridized at 4 °C overnight and – following 3 × 10 min washing in TBST20 – secondary antibodies (Supplementary Table 1) for 1 h at room temperature.

For western blotting of human proteins, infected cells were harvested in sample loading buffer (500 µl per well; six-well format), transferred to 1.5 ml reaction tubes, boiled for 5 min at 95 °C and 20 µl per lane were loaded onto a 10% PAA gel for SDS-PAGE as above. After blotting and blocking (as above), the membrane was probed with the respective primary antibody at 4 °C overnight and—upon washing (as above)—with the secondary antibody for 1 h at room temperature (a full list with information about all antibodies and sera used is given in Supplementary Table 1).

After three additional washing steps for each 10 min in TBST20, blots were developed using western lightning solution (Perkin Elmer) in a Fuji LAS-4000. In Fig. 3e, intensities of protein bands were quantified using the AIDA software (Raytest, Germany) and normalized to GroEL levels.

In vitro SPI-1 to SPI-2 switch assay. To mimic the early stages of the infection of a host cell *in vitro*, the indicated *Salmonella* strains were grown in LB overnight, diluted 1:100 in LB and grown to an OD₆₀₀ of 2.0 (that is, a condition under which SPI-1 is highly induced^{4,11}), washed twice with PBS and once with SPI-2 medium²⁸ at room temperature, diluted 1:50 in pre-warmed SPI-2 medium (defined as *t*₀) and grown further in Erlenmeyer flasks at 37 °C for the indicated time periods. At the respective time points, samples were taken for RNA-seq, western blotting, and GFP fluorescence measurements.

Bacterial GFP reporter measurements (flow cytometry, plate reader). To measure the GFP intensity of reporter strains, bacteria were grown in LB in presence of Amp and Cm until an OD₆₀₀ of 2.0 was reached. *Salmonella* cells corresponding to 1 OD₆₀₀ were pelleted and fixed with 4% PFA. GFP fluorescence intensity was quantified for each 100,000 events by flow cytometry with the FACSCalibur instrument (BD Biosciences). Data were analysed using the Cyflogic software (CyFlo).

To monitor SPI-2 activation in real time, a transcriptional *gfp* reporter was constructed by inserting the SPI-2-dependent *ssaG* promoter into plasmid pAS0093 via *AatII*/*NheI* sites as previously described⁸. The resulting plasmid pYC104 was co-transformed with either the pBAD-ctrl. or pBAD-PinT plasmid into the indicated strain backgrounds. The resulting strains were grown overnight in LB (+Amp +Cm) and then diluted 1:100 and further grown in the same medium to an OD₆₀₀ of 2.0. A volume of 1 ml of the culture was pelleted and the collected cells shifted to SPI-2 medium²⁸ (defined as *t*₀) as described above, except that the growth experiment was conducted in 96-well plates (Nunc Microwell 96F, Thermo Scientific). After measuring the OD₆₀₀ and GFP intensity at *t*₀, L-arabinose was added to each well to final concentration of 0.2% for sRNA induction and bacteria were grown for 20 h at 37 °C (with shaking) with measurements of both the OD₅₉₅ and GFP fluorescence in 10 min intervals using the Infinite F200 PRO plate reader (Tecan).

Enzyme-linked immunosorbent assay (ELISA). HeLa-S3 cells were infected with wild-type *Salmonella*, Δ *pinT* or *pinT*⁺ mutant strains at an m.o.i. of 5 as described above. Culture supernatant samples were taken at 20 h p.i. and analysed using the ELISA kit for human CXCL8/IL-8 (R&D Systems).

Bioinformatic analyses. Code availability. In order to document the details and parameters of the (dual) RNA-seq data analyses and to make the biocomputational approaches reproducible for others, we implemented the workflows as Unix Shell scripts. These scripts are deposited at Zenodo (DOI: 10.5281/zenodo.34695, <https://zenodo.org/record/34695>). Please refer to Supplementary Table 1 for descriptions of the analyses.

Read processing and mapping. For all RNA-seq experiments listed in Supplementary Table 1, Illumina reads in FASTQ format were trimmed with a Phred quality score cut-off of 20 by the program *fastq_quality_trimmer* from FASTX toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 20 nt after adaptor- and poly(A)-trimming were discarded before the mapping. The reads were aligned to the *Salmonella enterica* SL1344 genome (NCBI RefSeq accession numbers: NC_016810.1, NC_017718.1, NC_017719.1, NC_017720.1) and—where applicable—the human (hg19 – GRCh37; retrieved from the 1000 Genomes Project⁶³), the mouse (GENCODE M2, GRCm38.p2), or the porcine genome sequence (ENSEMBL, Sscrofa10.2), in parallel. The mapping was performed using the READemption pipeline (version 0.3.5)⁶⁴ and the short read mapper segemehl and its remapper lack (version 0.2.0)⁶⁵ allowing for split reads⁶⁶. Mapped reads with an alignment accuracy <90% as well as cross-mapped reads, that is, reads which could be aligned equally well to both host and *Salmonella* reference sequences, were discarded. The resulting data were used for visualization (see for example, Fig. 1b and Extended Data Fig. 2b).

Reads of the high resolution time-course experiment (cDNA libraries numbers 27–77 in Supplementary Table 1) that were detected as cross-mapped by READemption (see above) were further inspected: their median percentage over the entire time-course was 0.25% with increased fractions for the later time points, implying that those reads are mainly contributed by *Salmonella* cells. We observed that the majority of the cross-mapped reads aligned to *Salmonella* rRNA or tRNA loci, while on the human side no gene class preference was observed (data not shown).

Differential gene expression analysis. For dual RNA-seq experiments (cDNA libraries 1–184, 215–256 in Supplementary Table 1) after mapping differential expression analysis was carried out separately for the host and the pathogen. Strand-specific gene-wise quantifications for each data subset were performed by READemption⁶⁴. Host transcript expression analyses are based on annotations from GENCODE (version 19)⁶⁷, NONCODE (version 4)⁶⁸ and miRBase (version 20)⁶⁹ after removing redundant entries. The annotation for *Salmonella* genes was retrieved from NCBI (under the above mentioned accession numbers) and manually extended with small RNA annotations^{4,70}. In either organism, multi-mapped reads were removed and only uniquely mapped reads were considered for the expression analysis. Differential gene expression analyses were performed with the edgeR package (version 3.10.2)⁷¹ using an upper-quartile normalization and a prior count of 1. Where needed (that is, to correct for batch effects in the comparisons between wild-type and mutant infections; the comparisons displayed in Figs 3 and 4 and Extended Data Figs 5, 7–9), sequencing data were further normalized using the RUVs correction method⁷² with *k* = 3. For this purpose, we treated the samples time-point-wise to remove unwanted nuisance factors. At each time point our covariate of interest was the *pinT* status of the infecting bacterium. This is constant within replicate blocks, which are used for the RUVs correction. Host or bacterial genes with at least 10 uniquely mapped reads in three replicates were considered detected. Genes with an adjusted *P* value < 0.05 were considered differentially expressed. Differential expression analysis for conventional (bacteria only) RNA-seq experiments (cDNA libraries numbers 185–214 in Supplementary Table 1) was done similarly, except that a cut-off of ≥50 uniquely mapped reads was used as a detection threshold.

Read coverage plots. Based on the obtained BAM files, coverage files in wiggle format were generated by READemption⁶⁴ in a strand-specific manner and split by organism. In each case, coverage files are based on uniquely mapped reads and normalized by the total number of uniquely aligned reads per organism. For Fig. 4e, wiggle files were visualized using the Integrated Genome Browser (version 8.4.4)⁷³.

Pathway enrichment and further analyses for *Salmonella*. A database of pathways, regulons, and genomic islands was constructed using information obtained from the KEGG database⁷⁴ (organism code *sey*), the SL1344 genome annotation⁷⁰, and relevant literature sources (see Supplementary Table 1). Pearson correlation coefficients between changes in PinT expression and changes in expression of each gene within each regulon over the time-course of wild-type *Salmonella* infection (cDNA libraries number 27, 30, 33, 36, 39, 42, 44, 47, 50, 53, 56, 59, 61, 64, 67, 70, 73, 76 in Supplementary Table 1) were plotted in Fig. 2d.

To assess enrichment of differentially expressed transcripts in pathways in the comparative infection experiments (cDNA libraries numbers 27–77 and 152–184 in Supplementary Table 1) and the *in vitro* assay (cDNA libraries numbers 185–202 in Supplementary Table 1), gene set enrichment analysis (GSEA; version 2.1.0) was run on the log₂ fold changes reported by edgeR. The GSEA was performed in ranked list mode (with statistic classic) and gene sets containing less than 15 or more than 100 entries were excluded. Extended Data Fig. 5a reports all pathways significant at an FDR-corrected *P* value of at most 0.05 in at least one time point.

Host pathway analysis and inspection of alternative isoform expression. Host pathway enrichment studies were performed consistently with bacterial analyses using GSEA on human pathways available in the KEGG database (downloaded

January 22, 2014) using the same settings described above. Pathways with an adjusted P value ≤ 0.05 were considered to be significantly modulated. Data visualization for Extended Data Fig. 8a was produced using the Bioconductor package Pathview⁷⁵. Genes displayed in Fig. 1d, that is, genes whose transcription is known or predicted to be regulated by the binding of nuclear factor κ B (NF- κ B) to their promoter or genes whose products have been shown to promote an NF- κ B response, were retrieved from the GeneCards⁷⁶ and Boston University Biology (<http://www.bu.edu/nf-kb/gene-resources/target-gene>) databases or refs 77, 78. STAT3 target genes denoted in Fig. 4b were retrieved from ref. 79.

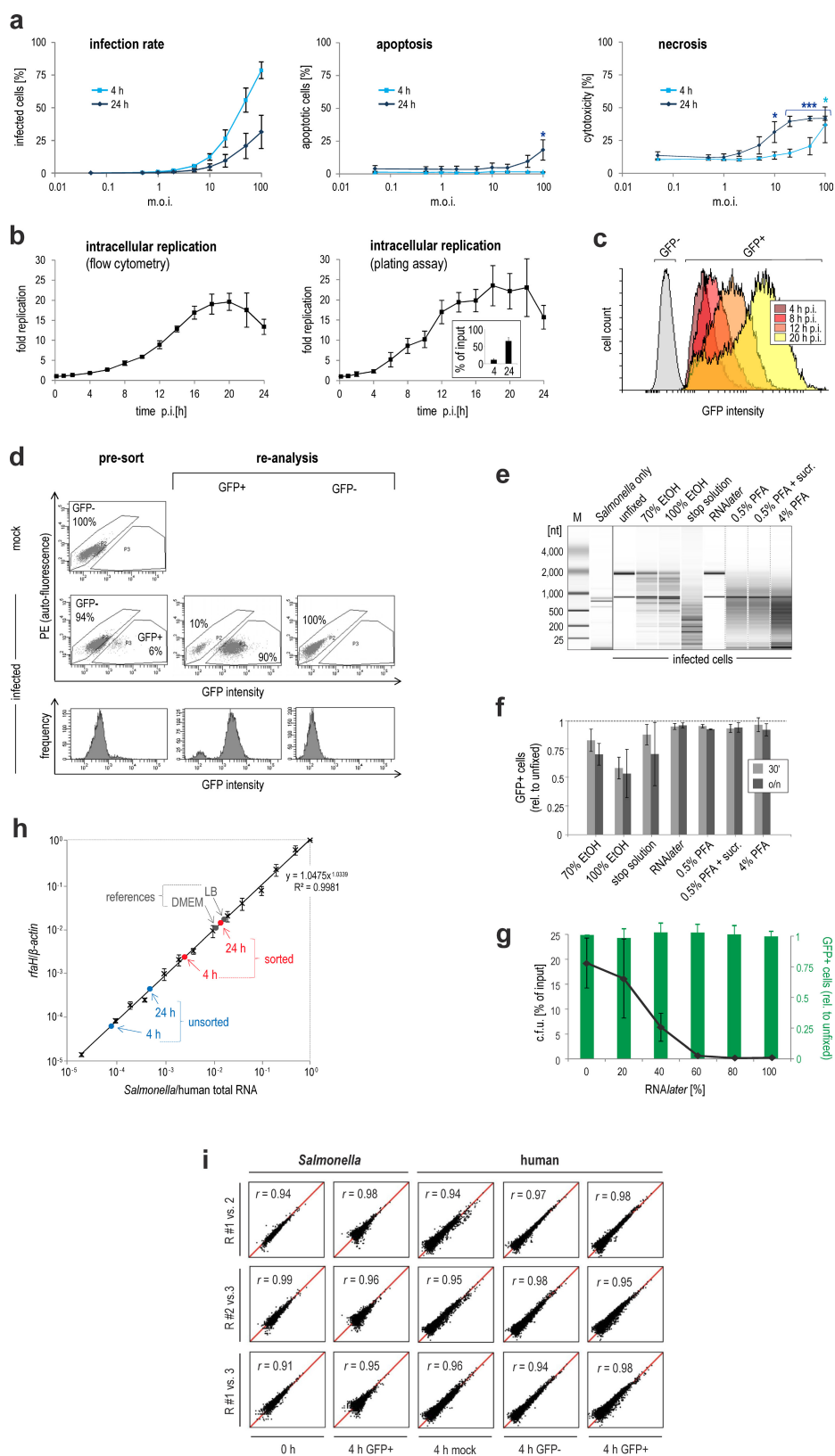
We used Cufflinks/Cuffdiff (version 2.2.1)^{80,81} to test for differentially expressed isoforms in the high-resolution, comparative dual RNA-seq time-course data set (cDNA libraries number 27–77 in Supplementary Table 1). In a first step, we used Cufflinks to quantify transcript isoforms in the mapped read data. Afterwards, all transcript annotations were merged using Cuffmerge and differentially expressed isoforms were called using Cuffdiff.

Interspecies correlation analysis of pathogen and host transcriptome changes. To identify bacterial and human genes with similar expression kinetics across the time-course of the infection of HeLa-S3 cells (cDNA libraries number 27–77 in Supplementary Table 1), we used RUVs-corrected, abundance-filtered and normalized read counts (see above). Absolute counts were then transformed into standard z-scores for each gene over all considered samples as follows: for each gene, the z-score was calculated as the absolute read count minus the mean read count over all samples, divided by the standard deviation of all counts over all samples. Genes with a standard deviation < 2 were excluded from further analysis. Pearson correlation coefficients were calculated between all remaining bacterial genes and all remaining human genes, and P values were calculated using the function `cor.test` in R. To account for a possible temporal delay between *Salmonella* expression changes and effect manifestation in the host cell, a time-shift was allowed. This means the expression of *Salmonella* genes at each time point was compared to host expression at the subsequent time point. Human genes were considered to be correlated with a bacterial gene if they had a P value of less than 10^{-4} and a Pearson's r greater than 0.65. This resulted in a total of 751 clusters of human genes showing correlation in expression with a bacterial gene, approximately half of which (see Supplementary Table 1) had at least one enriched GO term associated with them (adjusted P value < 0.05) as tested using the software tool Ontologizer 2.0 (build: 20100310-351)⁸² with the gene ontology definition obtained from the Gene Ontology Consortium (data-version: releases/2015-09-26) and the Universal Protein Resource (UniProt) gene annotation (generated: 2015-09-14).

To account for the possibility that multiple bacterial genes might be associated with a human gene cluster a correlation analysis was performed for all against all bacterial genes as described above, with the only exception that no time-shift was allowed. For this, we focused on seventeen gene clusters that were built on bacterial genes encoding for secretion-associated gene products (according to UniProt; see Supplementary Table 1). Detailed inspection of these clusters revealed the one depicted in Fig. 4b (centred on the bacterial SPI-2 gene *sseC*) which contained many further (bacterial and human) genes with pronounced PinT-dependent expression changes – that is, genes that showed differential expression between wild-type and Δ pinT infection at several time points p.i.

Statistics. In all RNA-seq-based analyses, transcript expression changes that were associated with an adjusted P value < 0.05 (reported by edgeR) were considered significantly differentially expressed. For Fig. 3b, a Monte Carlo permutation test was performed on the median fold change of genes in the SPI-2 regulon, using 10^5 randomly selected gene sets of the same size. This indicated the significant de-repression ($P < 0.05$) of the SPI-2 regulon in the absence of PinT at 2 and 8 h after the infection of HeLa cells, at 2, 6 and 16 h after the infection of 3D4/31 cells, and in the *in vitro* assay. Tests for the evaluation of increased host cell death in Extended Data Fig. 1a were performed using a one-tailed Student's t -test. $*P$ values ≤ 0.05 were considered significant and $***P$ values ≤ 0.001 were considered very significant. The significance of gene activation in qRT-PCR results in Fig. 4c and Extended Data Figs 5b, c and 7c, d or the ELISA assay in Extended Data Fig. 7e was assessed using a one-tailed Mann–Whitney U -test. The significance of differences in intracellular replication between the Δ pinT strain and wild-type *Salmonella* (Extended Data Fig. 4d) was evaluated using a two-tailed Mann–Whitney U -test.

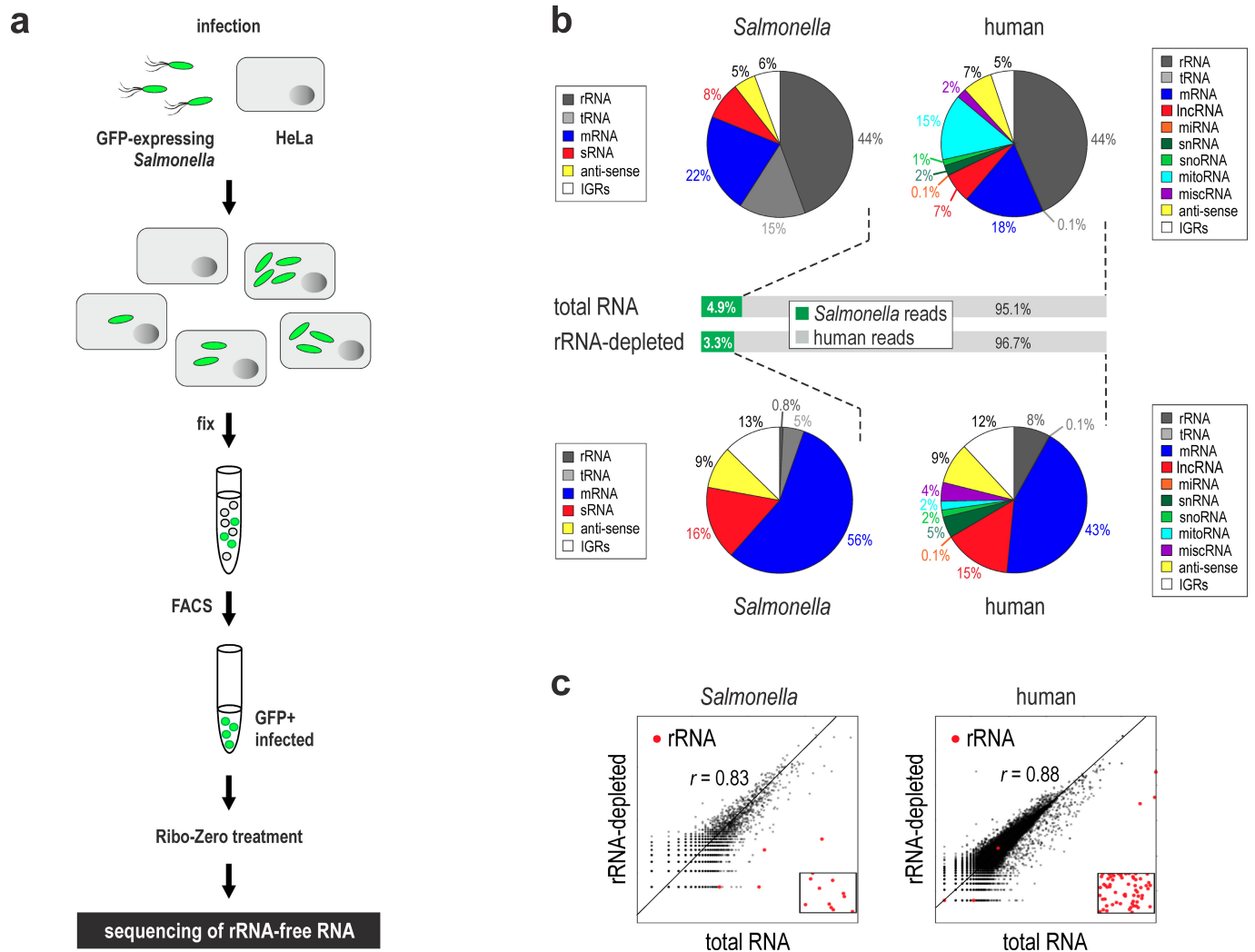
51. Papenfert, K. *et al.* Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol. Microbiol.* **74**, 139–158 (2009).
52. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
53. Sternberg, N. L. & Maurer, R. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. *Methods Enzymol.* **204**, 18–43 (1991).
54. Berschneider, H. M. Development of normal cultured small intestinal epithelial cell lines which transport Na and Cl. *Gastroenterology* **96**:A41 (1989).
55. Weingartl, H. M., Sabara, M., Pasick, J., van Moorleghem, E. & Babiuk, L. Continuous porcine cell lines developed from alveolar macrophages: partial characterization and virus susceptibility. *J. Virol. Methods* **104**, 203–216 (2002).
56. Schulte, L. N., Eulalio, A., Mollenkopf, H. J., Reinhardt, R. & Vogel, J. Analysis of the host microRNA response to *Salmonella* uncovers the control of major cytokines by the let-7 family. *EMBO J.* **30**, 1977–1989 (2011).
57. Eriksson, S., Lucchini, S., Thompson, A., Rhen, M. & Hinton, J. C. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol. Microbiol.* **47**, 103–118 (2003).
58. Hamatani, K. *et al.* Improved RT-PCR amplification for molecular analyses with long-term preserved formalin-fixed, paraffin-embedded tissue specimens. *J. Histochem. Cytochem.* **54**, 773–780 (2006).
59. Kuramochi, H. *et al.* Vascular endothelial growth factor messenger RNA expression level is preserved in liver metastases compared with corresponding primary colorectal cancer. *Clin. Cancer Res.* **12**, 29–33 (2006).
60. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **25**, 402–408 (2001).
61. Papenfert, K. *et al.* σ^E -dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol. Microbiol.* **62**, 1674–1688 (2006).
62. Urban, J. H. & Vogel, J. Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res.* **35**, 1018–1037 (2007).
63. The 1000 Genomes Project Consortium An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
64. Förstner, K. U., Vogel, J. & Sharma, C. M. READemption—a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* **30**, 3421–3423 (2014).
65. Otto, C., Stadler, P. F. & Hoffmann, S. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* **30**, 1837–1843 (2014).
66. Hoffmann, S. *et al.* A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. *Genome Biol.* **15**, R34 (2014).
67. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
68. Xie, C. *et al.* NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.* **42**, D98–D103 (2014).
69. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
70. Kroger, C. *et al.* The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc. Natl Acad. Sci. USA* **109**, E1277–E1286 (2012).
71. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
72. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnol.* **32**, 896–902 (2014).
73. Nicol, J. W., Helt, G. A., Blanchard, S. G., Raja, A. & Loraine, A. E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**, 2730–2731 (2009).
74. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205 (2014).
75. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831 (2013).
76. Rebhan, M., Chalifa-Caspi, P., Prilusky, J. & Lancet, D. GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.* **13**, 163 (1997).
77. Choi, Y. S. *et al.* Nuclear IL-33 is a transcriptional regulator of NF- κ B p65 and induces endothelial cell activation. *Biochem. Biophys. Res. Commun.* **421**, 305–311 (2012).
78. Oida, Y. *et al.* Inhibition of nuclear factor- κ B augments antitumor activity of adenovirus-mediated melanoma differentiation-associated gene-7 against lung cancer cells via mitogen-activated protein kinase kinase kinase 1 activation. *Mol. Cancer Ther.* **6**, 1440–1449 (2007).
79. Daur, D. J. *et al.* Stat3 regulates genes common to both wound healing and cancer. *Oncogene* **24**, 3397–3408 (2005).
80. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
81. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31**, 46–53 (2013).
82. Bauer, S., Grossmann, S., Vingron, M. & Robinson, P. N. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24**, 1650–1651 (2008).
83. Soncini, F. C., Vescovi, E. G. & Groisman, E. A. Transcriptional autoregulation of the *Salmonella typhimurium* *phoPQ* operon. *J. Bacteriol.* **177**, 4364–4371 (1995).
84. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–1517 (2004).



Extended Data Figure 1 | See next page for caption.

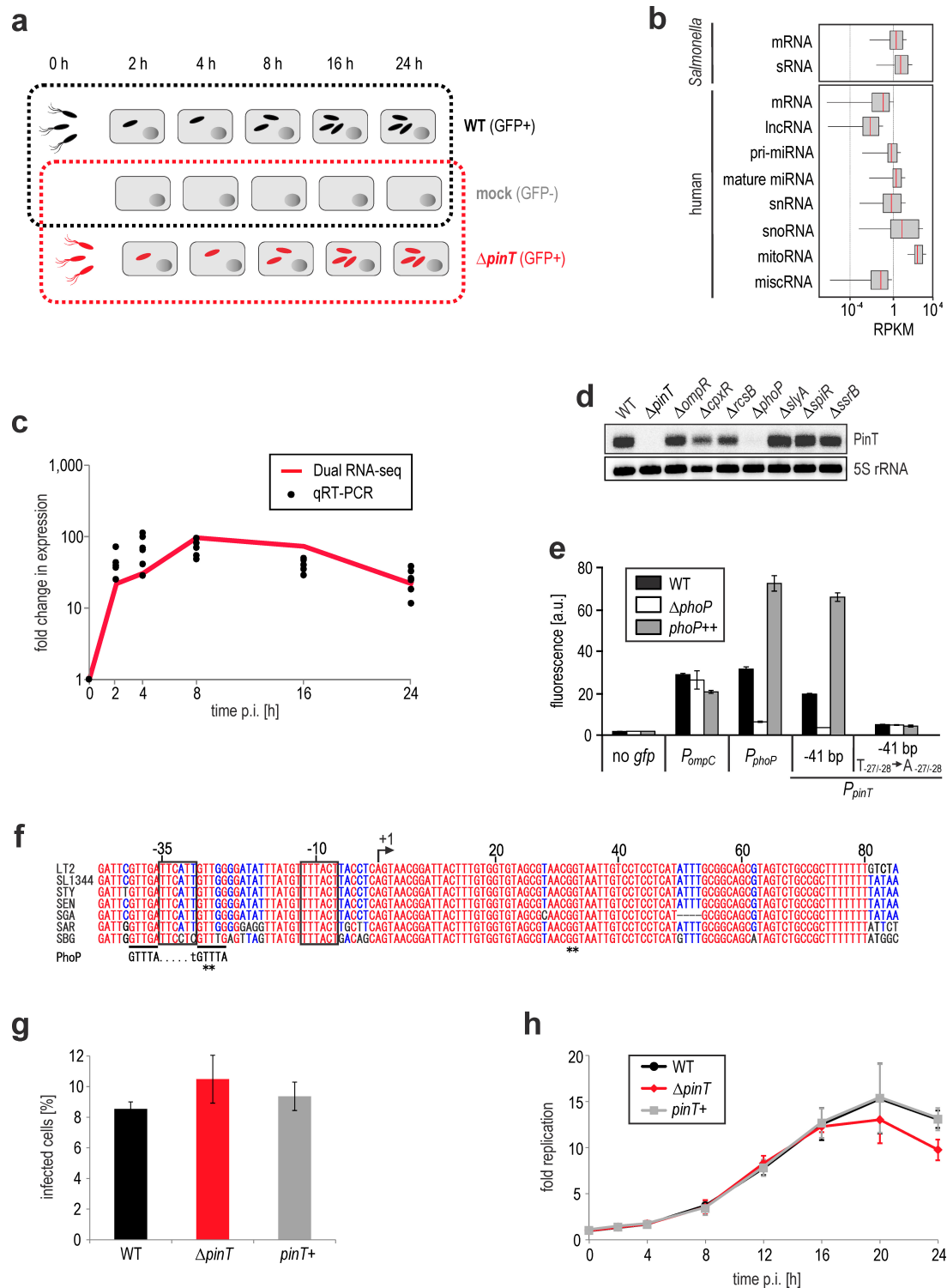
Extended Data Figure 1 | Establishment of the infection model with HeLa-S3 cells and constitutively GFP-expressing *Salmonella*. **a**, Rate of infected, apoptotic and cytotoxic HeLa-S3 cells over a range of different m.o.i.s. Left panel: infectivity increases with increasing bacterial doses. The discrepancy between the fractions of infected cells at 4 h and 24 h p.i. results from increasing levels of host cell death over time. Quantification of infectivity was achieved via flow cytometry (FACSCalibur, BD Biosciences) and the Cyflogic software (CyFlo) by gating for the GFP⁺ and GFP⁻ populations. Middle: apoptosis measurement by annexin V-APC (BD Pharmingen)/propidium iodide (PI) (Sigma) staining followed by flow cytometry using the MACSQuant Analyzer device (Miltenyi Biotec). APC-positive/PI-negative cells were considered apoptotic. Right: lactate dehydrogenase (LDH) release as a proxy for necrosis in infected HeLa-S3 cultures. The colorimetric product was quantified by measuring the absorbance at 490 nm. 100% host cell death was determined by treating the cell culture with lysis buffer before analysis. * $P \leq 0.05$; *** $P \leq 0.001$ (one-tailed Student's *t*-test). **b**, Intracellular replication of *Salmonella* inside HeLa-S3 (m.o.i. of 5). Left panel, flow cytometry-based quantification of the increase in GFP intensity per infected host cell over time (see also panel c and Methods). Right: c.f.u. counts. The inset illustrates the relative amount of intracellular bacteria at 4 or 24 h p.i., respectively, as compared to the input. Combination of infectivity (panel a) and c.f.u. data (panel b) allows for the calculation of the average number of intracellular bacteria per invaded cell at distinct time points: 4 h p.i.: ~10 bacteria; 24 h p.i.: ~75 bacteria. **c**, Representative overlay histogram of flow cytometry data exemplifying the increase in GFP intensity per infected HeLa-S3 cell over time. Plot was generated using the Flowing software (Turku Centre for Biotechnology, Finland). **d**, Representative FACS plots showing the

gating strategy for the separation of invaded and non-invaded HeLa-S3 cells. The signal detected in the phycoerythrin (PE) channel was used as a proxy for a cell's autofluorescence. The percentage values indicate the relative proportion of GFP⁺ and GFP⁻ cells before ('pre-sort') and after sorting ('re-analysis'). **e**, Capillary electrophoresis of total RNA samples from infected cells (4 h p.i.) that were fixed overnight using different reagents. Stop solution refers to 95% EtOH/5% phenol. Where indicated ('+sucr.') paraformaldehyde (PFA) was supplemented with 2% sucrose. The band pattern of pure *Salmonella* RNA is shown to the left. Note that in the infection samples bacterial rRNA bands are not visible due to the overwhelming host RNA background. For gel source data, see Supplementary Fig. 1. **f**, Influence of different preservatives on FACS-based recovery of invaded host cells through bleaching of the GFP signal. **g**, Increasing concentrations of RNAlater kill intracellular *Salmonella* (black line) but do not compromise detection of GFP fluorescence (green bars). **h**, Extrapolation of the relative representation of *Salmonella* and human transcriptomes in infection samples. A dilution series of separately isolated *Salmonella* to HeLa-S3 total RNA was set up and for each ratio, bacterial *rfaH* (relative to human *ACTB*) mRNA was quantified by qRT-PCR. The resulting trend-line equation was used to infer the percentage of the *Salmonella* transcriptome within mixed total RNA samples from infected HeLa-S3 cells at different time points and without (blue) or upon FACS-based enrichment for invaded cells (red). The position of medium control samples (LB, DMEM) is given in grey. **i**, Normalized read counts for all detected *Salmonella* or human genes at 4 h p.i. are plotted for 3 biological replicate experiments and the Pearson correlation coefficient (*r*) is given. Panels a, b, f, g, h show the mean \pm s.d. from each 3 biological replicates.



Extended Data Figure 2 | Establishment of an rRNA-depletion step for dual RNA-seq. **a**, Experimental workflow. **b**, Comparative mapping statistics of dual RNA-seq samples without (upper panel) or upon the joint depletion of bacterial and eukaryotic rRNA using the Ribo-Zero Gold (Epidemiology) kit (lower panel). **c**, Gene-wise correlation between read

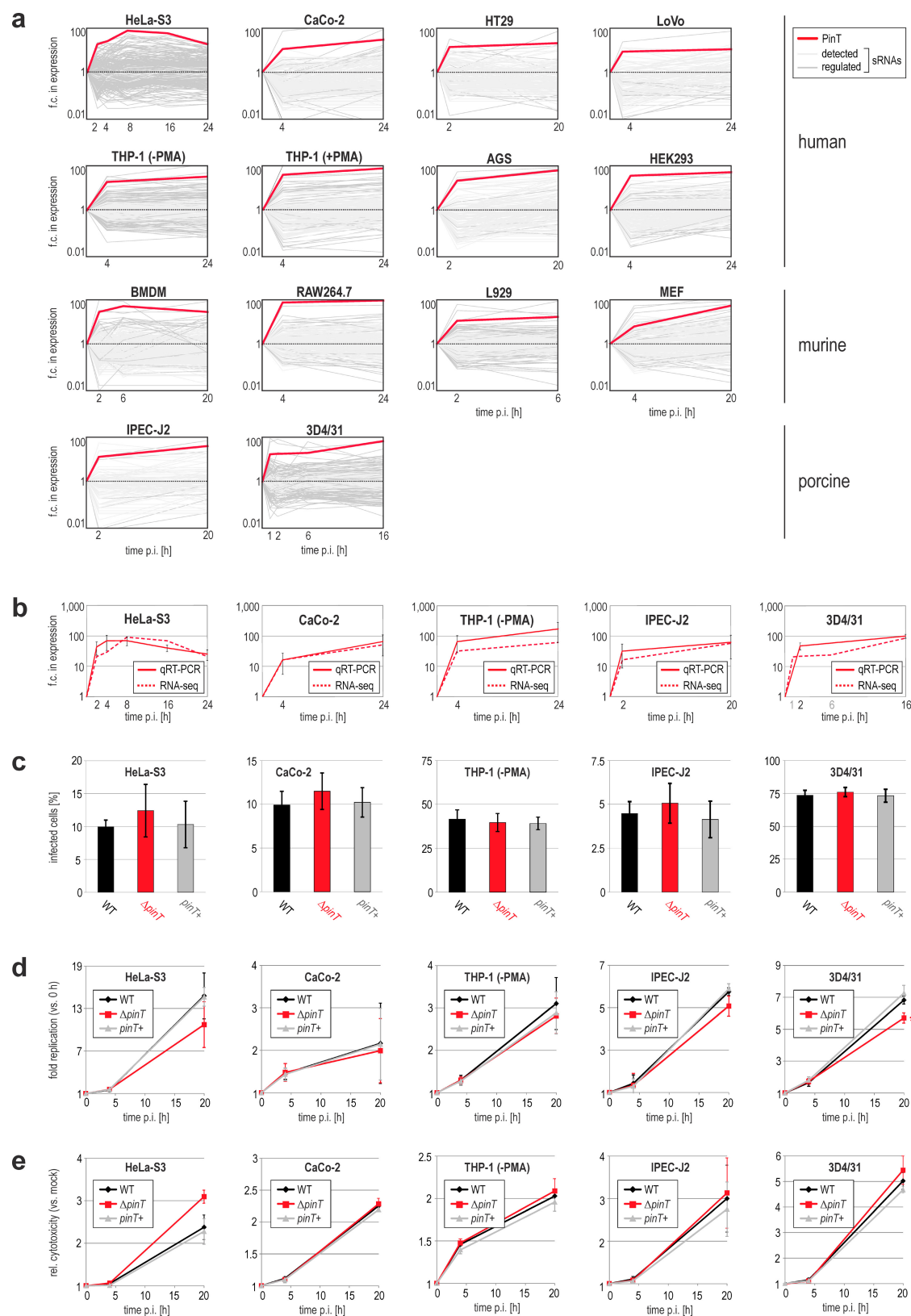
coverages without or upon rRNA removal for *Salmonella* (left) and human (right) data subsets. The Pearson's r is given. The red dots in squares represent the rRNA transcripts that had zero reads in the rRNA-depleted sample.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | PinT is induced during infection via PhoP binding to its promoter region. **a**, Scheme of the comparative high-resolution time-course analysed by dual RNA-seq. For both the *Salmonella* strains, five individual time points post-invasion of HeLa-S3 cells were sampled and enriched for the fraction of invaded (GFP⁺) host cells. Mock-treated cells were used as a host, and extracellular *Salmonella* in DMEM medium (0 h) as a bacterial reference control. Together this resulted in 17 different conditions which were sampled as biological triplicates. **b**, Average RPKM distribution over individual *Salmonella* or human transcript classes from the wild-type infection time-course. **c**, PinT activation during invasion of HeLa-S3 cells as revealed by dual RNA-seq (red graph) can be reproduced by qRT-PCR measurements (black dots; each dot represents a single out of 4 (2; 8; 16 h), 5 (4 h) or 6 biological replicate experiments (24 h)). Normalization was achieved using the constitutively expressed *gfp* mRNA as a reference. **d**, Northern blot detection of PinT in the *Salmonella* wild-type and various mutant backgrounds in which the indicated global regulators were deleted. For gel source data, see Supplementary Fig. 1. **e**, Mutational analysis identifies PhoP as a direct transcriptional activator of the *pinT* promoter. A transcriptional *gfp* fusion construct containing the *pinT* upstream promoter region (−41 bp to +5 bp) was analysed in the wild-type,

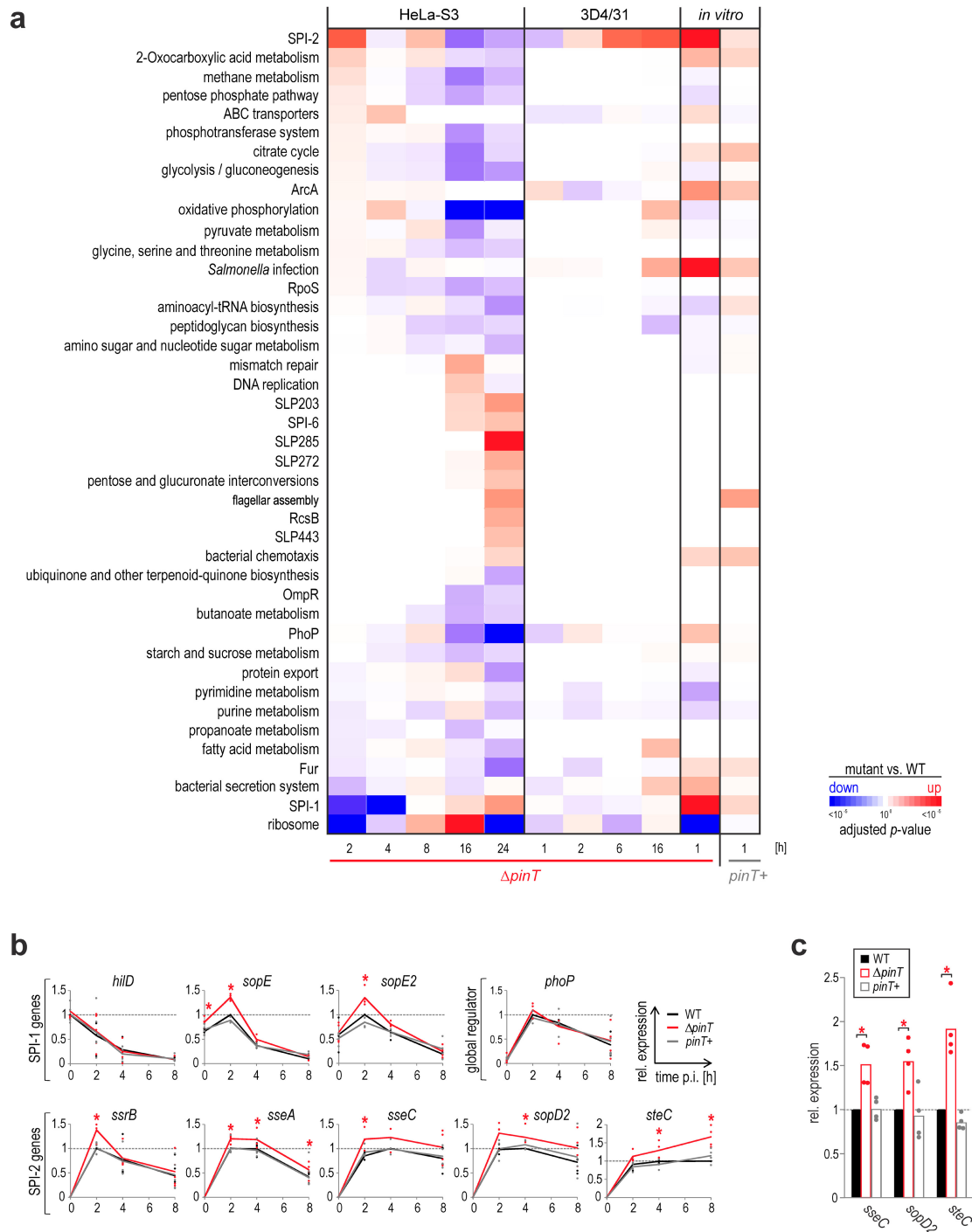
phoP deletion or *phoP* complementation backgrounds. A non-fluorescent ('no *gfp*') or unrelated *ompC* promoter reporter served as negative controls and a *phoP* promoter reporter as a positive control (as PhoP is known to auto-regulate its own expression⁸³). Two-nucleotide exchanges (T_{−27/−28}→A_{−27/−28}) in the predicted PhoP binding site (see alignment in panel **f**) are sufficient to abrogate PhoP responsiveness of PinT expression. Error bars indicate the s.d. from the mean from biological triplicates. **f**, Sequence alignment shows the conservation of PinT sRNA within the genus *Salmonella*. "STY": *S. Typhi*, "SEN": *S. Enteritidis*, "SGA": *S. Gallinarum*, "SAR": *S. arizonae*, "SBG": *S. bongori*. Perfectly conserved ribonucleobases are labelled in red, less conserved bases are shown in blue. The numbers indicate the position relative to the 5' end of PinT (+1 position). Black lines and sequence motif below the alignment highlight a PhoP binding site (the asterisks denote thymines that were converted into adenines for mutational analysis in panel **e**). Asterisks below the seed sequence (position ~30–40) mark two guanines that were mutated to cytosines in Fig. 3c and Extended Data Fig. 6. **g, h**, Infection rate (**g**) and intracellular replication kinetics (**h**) of the indicated *Salmonella* strains in HeLa-S3 cells (m.o.i. of 5). Data are derived from flow cytometry measurements as described for Extended Data Fig. 1a, b, and refer to the mean ± s.d. from 3 biological replicates.



Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | PinT is strongly induced upon the invasion of diverse host cell types while its deletion only has slight effects in cell culture models. **a**, The given cell types were infected with wild-type *Salmonella* for the indicated time periods and total or rRNA-depleted RNA (as indicated in Supplementary Table 1) was sequenced. THP-1 cells either were differentiated by treating them with phorbol myristate acetate before infection ('+PMA') or were kept monocytic ('-PMA'). For all but HeLa-S3 cells and porcine cell types infection was established at an m.o.i. of 10. IPEC-J2 and 3D4/31 cells were infected at an m.o.i. of 20 and HeLa-S3 at an m.o.i. of 5. Shown are detected (≥ 10 reads in each replicate; light grey) and regulated sRNAs (adjusted P value < 0.05 ; dark grey). The data was derived from 3 biological replicates for HeLa-S3 and 3D4/31 and 2 biological replicates for the other cell types. PinT expression (red line) was significantly upregulated in all cell types. **b**, qRT-PCR validation of the induction of PinT in five selected host cell types. qRT-PCR data from 3 biological replicates were drawn (solid lines). Normalization was against *gfp* mRNA. The data from HeLa-S3 is the same as in Extended Data Fig. 3c.

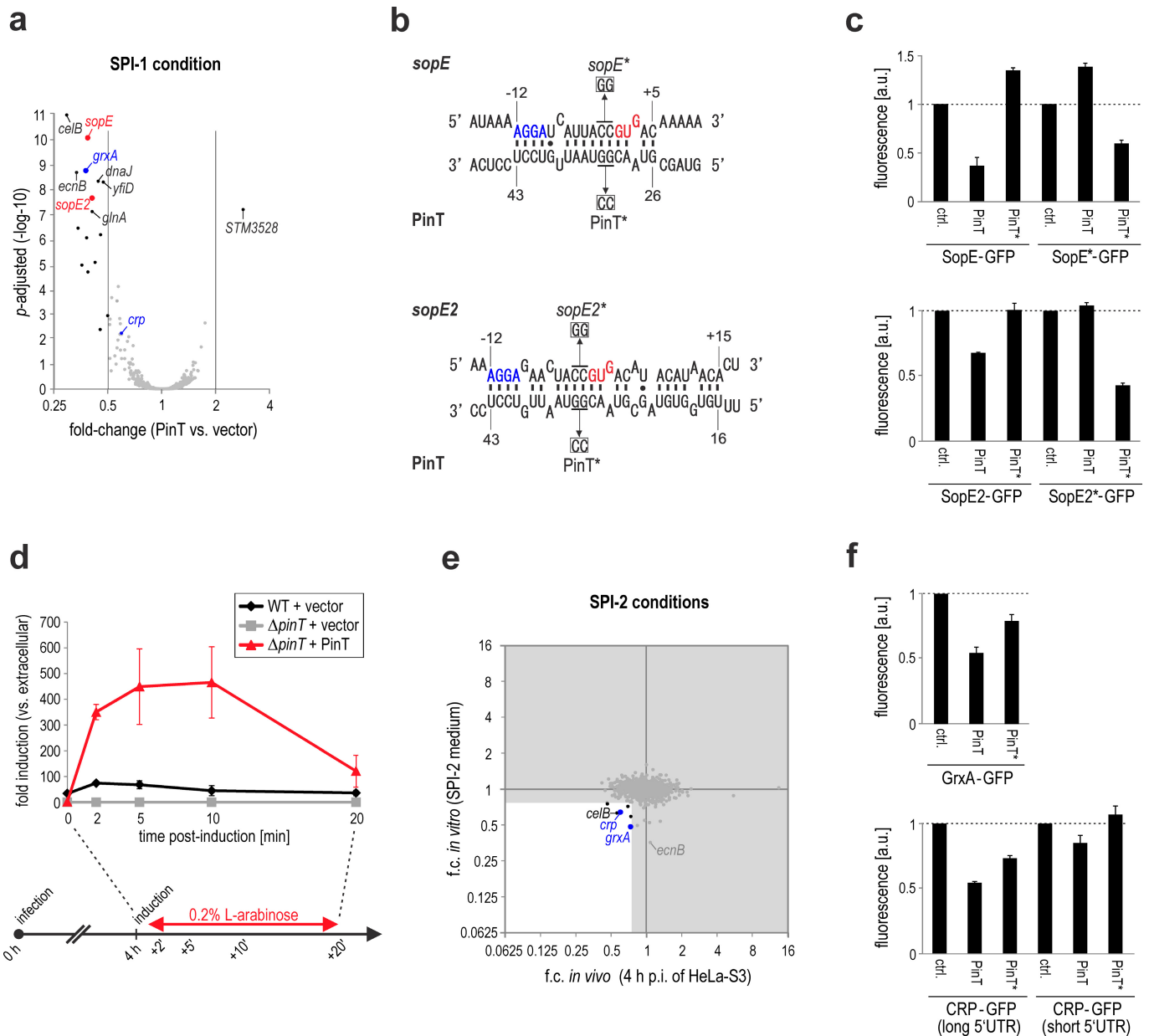
For comparison, in each case the RNA-seq-based expression data of PinT (as shown in panel **a**) are re-plotted (dashed curves). **c**, Invasion assays for wild-type, $\Delta pinT$ and *pinT*+ *Salmonella* with HeLa-S3 (m.o.i. 5), CaCo-2 and undifferentiated THP-1 (both m.o.i. 10), as well as IPEC-J2 and 3D4/31 (both m.o.i. 20). In all cases, the invasion rate was profiled 10 min p.i. by flow cytometry. **d**, Intracellular replication kinetics for the same strains and host cell types. The increase in GFP intensity of infected cells over time was monitored by flow cytometry and expressed as fold change compared to the t_0 time point (see Methods). The asterisk denotes a significantly different replication rate between the wild-type and $\Delta pinT$ strain ($P < 0.05$; two-tailed Mann-Whitney U -test). **e**, Host cell cytotoxicity measurements for the same host cell types upon infection with the indicated *Salmonella* strains. At the respective time points p.i., LDH activity in the supernatant of the infected cultures was quantified and the increase over time was with respect to the LDH activity measured in supernatants of mock-infected cells. The data in panels **c-e** represent the mean \pm s.d. from 3 biological replicates.



Extended Data Figure 5 | *Salmonella* PinT sRNA represses SPI-2 expression during the infection of HeLa-S3 cells and pig macrophages.

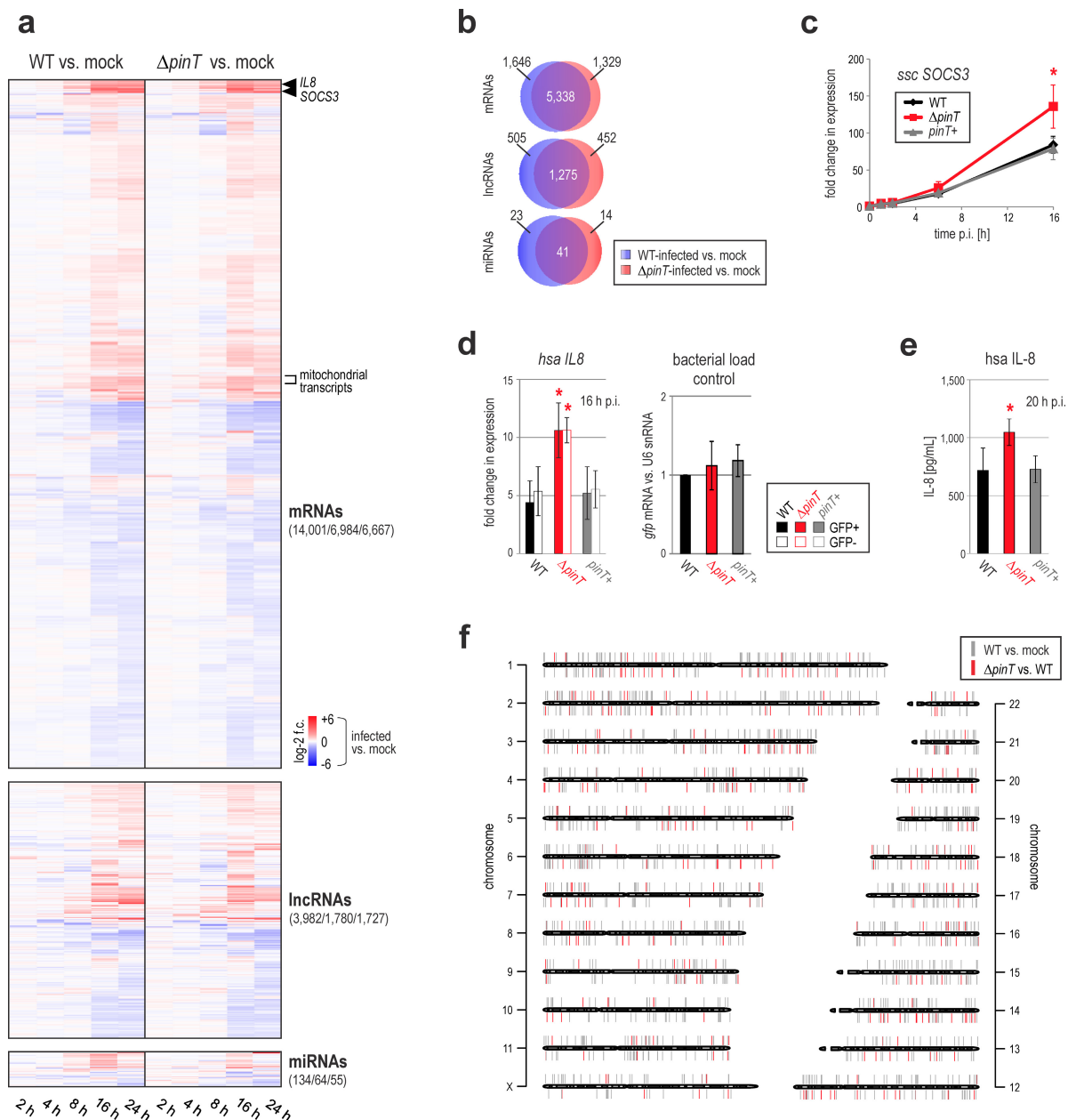
a, The heat map shows the result from gene set enrichment analyses of *Salmonella* gene expression data from the comparative dual RNA-seq time-course experiments with HeLa-S3 cells and porcine 3D4/31 macrophages, and from the comparative RNA-seq experiment of *Salmonella* grown for 1 h under SPI-2-inducing *in vitro* conditions. It reveals the de-repression of SPI-2 genes in the absence of PinT ($\Delta pinT$) at several time points—a specific effect as SPI-2 expression reverts to wild-type levels upon *trans*-complementation of PinT (*pinT*+) in the *in vitro* assay. **b**, qRT-PCR measurements during the early stages of HeLa-S3

infection validates the de-repression of *Salmonella* SPI-2 genes in the $\Delta pinT$ background (red) as compared to both wild-type (black) and *pinT*+ (grey) strains. Normalization was performed using *gfp* mRNA. Dots represent individual biological replicate experiments (five for *hliD*, *sseA*, *ssrB*; four for the other mRNAs) and the solid lines indicate their mean. **c**, qRT-PCR validation of de-repressed SPI-2 genes 6 h after the invasion of pig macrophages. Porcine *GAPDH* mRNA was used for normalization. The data represent the results from biological triplicate measurements. The asterisks in panels **b** and **c** denote significantly increased transcript levels in $\Delta pinT$ compared to wild-type *Salmonella* ($P < 0.05$; one-tailed Mann-Whitney *U*-test).



Extended Data Figure 6 | PinT directly targets *Salmonella* *sopE/E2*, *grxA* and *crp* mRNAs. **a**, Volcano plot showing *Salmonella* mRNA levels at 5 min after the pulse-expression of PinT under SPI-1-inducing conditions (LB medium, OD₆₀₀ of 2.0). The data are derived from two biological replicates. Candidates that were confirmed to be directly targeted by PinT are coloured. Red: targets regulated predominantly under SPI-1 conditions; blue: targets regulated (also) under SPI-2 conditions (see panel **e**). For the full list of changes in gene expression after the PinT pulse see Supplementary Table 1. **b**, RNA duplex formation between PinT and the *sopE* and *sopE2* leaders as predicted by the RNA-hybrid program⁸⁴. Point mutations introduced for compensatory base-pair exchange experiments are indicated. The ribosome binding site and start codon are marked in blue or red, respectively. **c**, Validation of the base-pair interactions as shown in panel **b** using translational *sopE::gfp* and *sopE2::gfp* reporter gene fusions by compensatory base-pair exchanges. *Salmonella* strains containing both a *gfp* reporter plasmid and an sRNA overexpression vector were grown overnight in LB and analysed by flow cytometry. The error bars indicate s.d. from the mean from biological triplicates. **d**, In vivo pulse-expression establishment. GFP-expressing

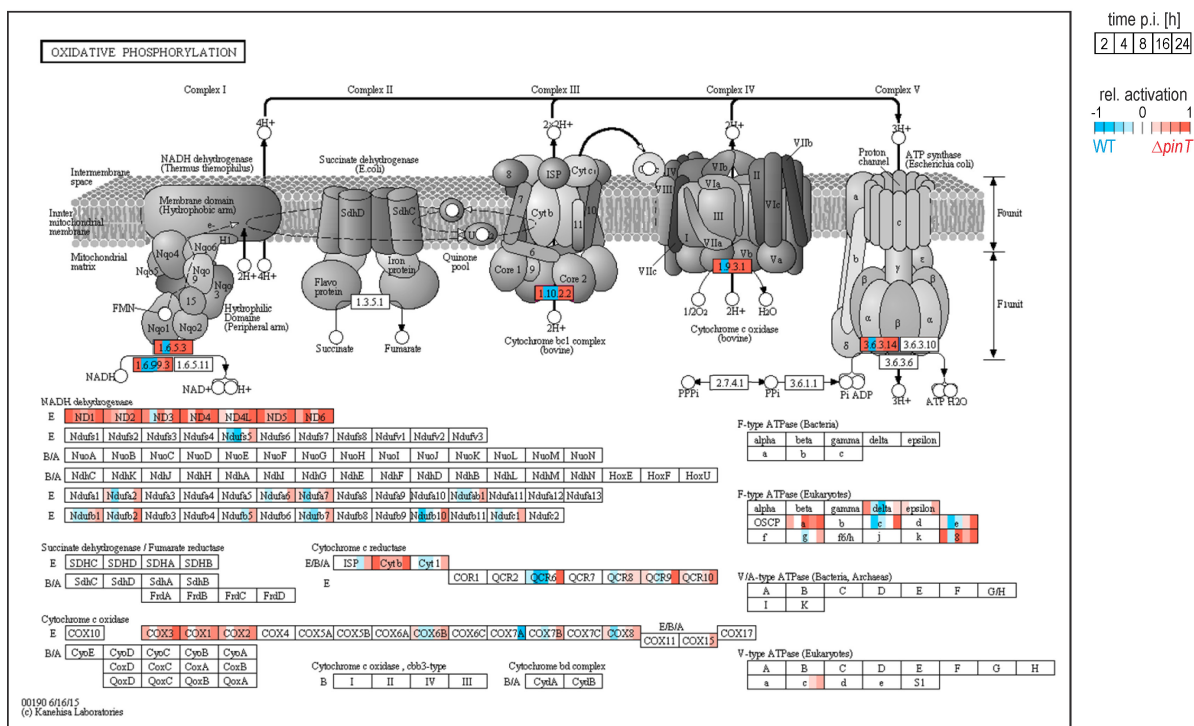
Salmonella strains harbouring PinT sRNA under an λ -arabinose-inducible promoter on a plasmid or corresponding control strains, respectively, were used to infect HeLa-S3 cells. At 4 h p.i., λ -arabinose was added to the cell medium. Samples were taken over a time-course of 20 min after the pulse and enriched for *Salmonella* transcripts (see Methods section). PinT sRNA levels were measured by qRT-PCR in the resulting RNA samples and are plotted (mean \pm s.d. from technical triplicates). **e**, Pulse-expression of PinT under *in vivo* (-like) conditions (the full data are in Supplementary Table 1). PinT was transiently overexpressed under SPI-2-inducing conditions *in vitro* or 4 h after HeLa-S3 infection (see panel **d**). In either case, 5 min after induction total RNA samples were taken and analysed by RNA-seq (each two biological replicates). Axes represent fold changes in mRNA abundance between strains harbouring the empty and the PinT-containing plasmid. The two targets validated in panel **f** are labelled in blue. The *celB* (cellobiose-specific permease IIC component) and *ecnB* (entericidin B precursor) mRNAs might be further targets of PinT (see also panel **a**), but were not followed up here. **f**, Validation of direct targeting of *grxA* and *crp* mRNAs by the seed region of PinT using translational *grxA::gfp* and *crp::gfp* reporter gene fusions as described for panel **c**.



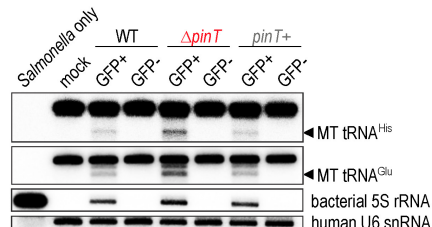
Extended Data Figure 7 | Host expression data for the comparative infection with wild-type or $\Delta pinT$ *Salmonella*. **a**, Heat maps showing differentially expressed mRNAs (upper panels), lncRNAs (middle) or miRNAs (lower) between the infection of HeLa-S3 cells with either one of the two indicated *Salmonella* strains and mock-infected controls. Plotted are all genes that were significantly differentially expressed (adjusted P value < 0.05 ; 3 biological replicates) between the indicated conditions for at least one time point during infection. Numbers to the right refer to detected transcripts and transcripts differentially expressed after wild-type infection vs mock, or after $\Delta pinT$ infection vs mock, respectively. **b**, PinT affects the expression of coding and noncoding transcripts of the human host. Venn diagrams indicate HeLa-S3 transcripts commonly or specifically regulated compared to mock for the respective infection strain. **c**, Infection with $\Delta pinT$ *Salmonella* leads to increased SOCS3 expression in porcine macrophages. qRT-PCR data (mean \pm s.d.) from biological triplicate experiments of the infection of 3D4/31 cells with *Salmonella* wild-type, $\Delta pinT$, or $pinT^+$. Total RNA samples were taken at 6 h p.i. Porcine GAPDH mRNA was used for normalization. **d**, qRT-PCR measurement of human IL8 mRNA in total RNA samples isolated from HeLa-S3 cells at 16 h after either wild-type, $\Delta pinT$ or $pinT^+$ infection

and each sorted into the fractions of invaded (GFP⁺) and non-invaded (GFP⁻) cells. U6 snRNA was used for normalization. The data represent the mean \pm s.d. from 3 biological replicates. Note that differential gene expression was not caused by different bacterial loads as judged from the ratio of *gfp* mRNA to U6 snRNA ('bacterial load control'). **e**, Enzyme-linked immunosorbent assay for human IL-8 protein in supernatant samples from HeLa-S3 cells at 20 h after their infection with the indicated *Salmonella* strains. Data refer to the mean \pm s.d. from biological triplicate measurements. **f**, Karyogram plot displaying the individual human (female) chromosomes and the genomic position of differentially expressed lncRNA candidates. lncRNAs differentially regulated (adjusted P value < 0.05 ; 3 biological replicates) in response to wild-type infection compared to mock-treated HeLa-S3 cells are indicated as grey bars and candidates differentially expressed between wild-type- and $\Delta pinT$ -infected cells as red bars. The position of the bars relative to the respective chromosome indicates the direction of regulation (above the chromosome: upregulation; below: downregulation at the earliest time point when regulation was observed). Panels **c**–**e**, asterisks denote significantly different transcript (**c**, **d**) or protein (**e**) levels between wild-type- and $\Delta pinT$ -infected cells ($P < 0.05$; one-tailed Mann–Whitney U -test).

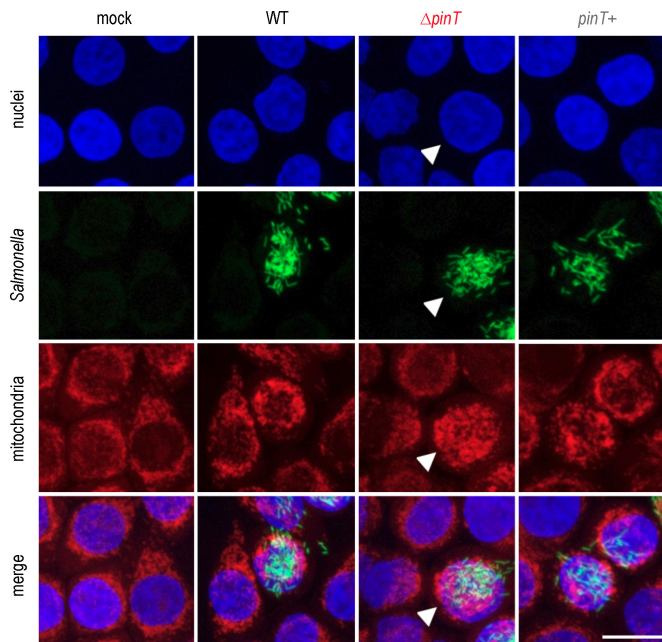
a



b

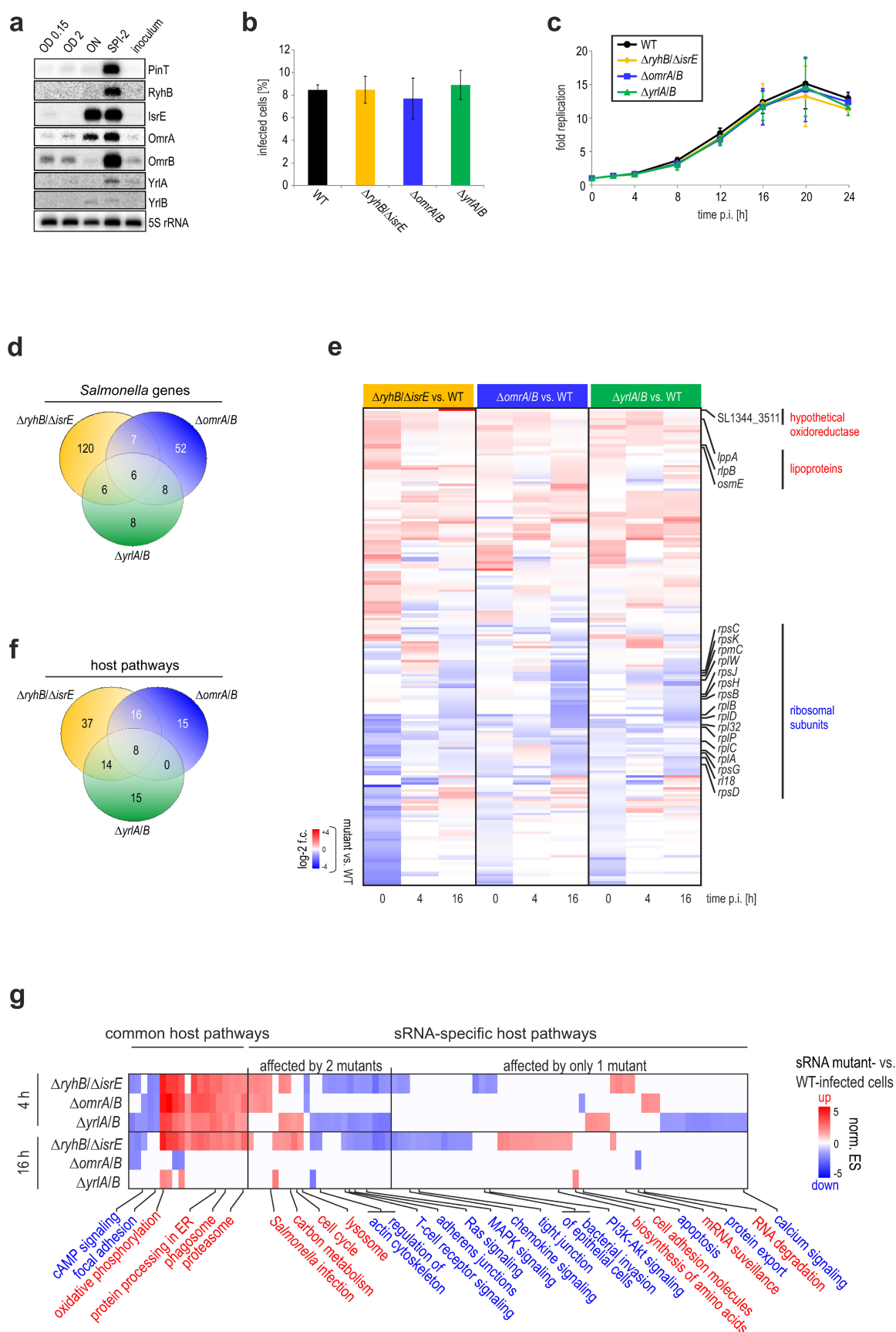


C



Extended Data Figure 8 | Impact of PinT on host mitochondria. a, KEGG Pathview representation of oxidative phosphorylation in mitochondria. Dual RNA-seq data of infected HeLa-S3 cells (3 biological replicates) were plotted on top of the pathway map. Individual boxes represent the different time points sampled. Coloured boxes: at the respective time point the given gene(-cluster) was hyper-activated (red) or suppressed (blue) in $\Delta pinT$ -infected compared to wild-type-infected cells. **b,** Northern blot detection of mitochondrial tRNAs (shown in Fig. 4e) in HeLa-S3 cells infected for 16h with the indicated *Salmonella* strains and sorted for GFP⁺ and GFP⁻

fractions. *Salmonella* 5S rRNA serves as a bacterial and human U6 snRNA as a host control. The 'Salmonella only' sample demonstrates specificity of the probes against the respective human (but not bacterial) tRNAs. For gel source data, see Supplementary Fig. 1. c, Elevated mitochondrial expression in response to $\Delta pinT$ infection is accompanied by the sub-cellular re-localization of mitochondria in invaded hosts. Mitochondria of infected HeLa-S3 cells were stained using the MitoTracker Orange dye (Life Technologies) and nuclei with Hoechst (Invitrogen). The scale bar indicates 15 μm . The white arrowhead marks a prominent cluster of re-localized mitochondria.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Comparative dual RNA-seq experiments with further sRNA mutant *Salmonella*. **a**, Northern blot detection of additional sRNAs induced upon host cell invasion (see Fig. 2a). 'Inoculum' refers to bacteria in DMEM. Gel source data in Supplementary Fig. 1. **b**, HeLa-S3 invasion efficiency of the three indicated sRNA double mutants and wild-type *Salmonella* (m.o.i. of 5). **c**, Intracellular replication kinetics of the same strains inside HeLa-S3 (m.o.i. of 5). Data in panels **b** and **c** are derived from flow cytometry measurements as described for Extended Data Fig. 1a, b, and refer to the mean \pm s.d. from 3 biological replicates. **d**, Dual RNA-seq data of the infection of HeLa-S3 cells with the indicated *Salmonella* strains at 0 h, 4 h, and 16 h p.i. The Venn diagram shows the number of significantly (adjusted P value < 0.05) differentially expressed *Salmonella* genes (combined for all three time points) between wild-type *Salmonella* and the respective sRNA double mutant strain. **e**, Heat map of bacterial mRNA and sRNA expression changes for $\Delta omrA/B$, $\Delta rylB/\Delta isrE$ and $\Delta yrlA/B$ as compared to wild-type

Salmonella at 0 h, 4 h and 16 h p.i. of HeLa-S3 cells. Plotted are all genes (except the respectively deleted sRNAs) that were significantly differentially expressed (adjusted P value < 0.05) for at least one of the indicated conditions. The gene for a putative oxidoreductase (*SL1344_3511*) was specifically upregulated in intracellular $\Delta rylB/\Delta isrE$ mutants compared to wild-type *Salmonella*. Many genes for ribosomal subunits were downregulated in all three mutant strains compared to the wild-type. **f**, Venn diagram for the number of commonly or specifically affected human pathways between *Salmonella* wild-type infection and that of the three mutant strains (further specified in panel **g**). **g**, Infection of HeLa-S3 with three sRNA double mutant strains affects distinct sets of host pathways as compared to wild-type infection. Shown are normalized enrichment scores (norm. ES) from a human gene set enrichment analysis (adjusted P value < 0.05). The dual RNA-seq data in panels **d–g** were derived from 3 biological replicate experiments.

Formation of new stellar populations from gas accreted by massive young star clusters

Chengyuan Li^{1,2,3}, Richard de Grijs^{1,4}, Licai Deng², Aaron M. Geller^{5,6}, Yu Xin², Yi Hu² & Claude-André Faucher-Giguère⁵

Stars in clusters are thought to form in a single burst from a common progenitor cloud of molecular gas. However, massive, old ‘globular’ clusters—those with ages greater than ten billion years and masses several hundred thousand times that of the Sun—often harbour multiple stellar populations^{1–4}, indicating that more than one star-forming event occurred during their lifetimes. Colliding stellar winds from late-stage, asymptotic-giant-branch stars^{5–7} are often suggested to be triggers of second-generation star formation. For this to occur⁸, the initial cluster masses need to be greater than a few million solar masses. Here we report observations of three massive relatively young star clusters (1–2 billion years old) in the Magellanic Clouds that show clear evidence of burst-like star formation that occurred a few hundred million years after their initial formation era. We show that such clusters could have accreted sufficient gas to form new stars if they had orbited in their host galaxies’ gaseous disks throughout the period between their initial formation and the more recent bursts of star formation. This process may eventually give rise to the ubiquitous multiple stellar populations in globular clusters.

The colour–magnitude diagrams of NGC 1783, NGC 1696 and NGC 411 are shown in Fig. 1. These stellar distributions—the observational counterparts of the theoretical Hertzsprung–Russell diagrams, which relate the stellar surface temperatures to their luminosities—have been field-star-decontaminated by careful application of statistical background-subtraction techniques (see Methods). Figure 1a shows that the majority of stars associated with NGC 1783 are well-described by an isochrone⁹—a theoretical ridge-line that describes stars with identical ages but covering a range of initial masses—characterized by an age t given by $\log t = 9.15$ (that is, $t = 1.4 \times 10^9$ yr; see Methods). However, one can also clearly discern two additional, bright stellar sequences, denoted ‘A’ and ‘B’, with younger ages of, respectively, $\log t = 8.65$ ($t = 450$ Myr) and $\log t = 8.95$ ($t = 890$ Myr). These two sequences appear to have chemical compositions that are similar to the cluster’s bulk stellar population (particularly in terms of the abundances of helium and heavier elements), given the absence of any clear differences in the observed ridge-line colours. Sequence A may also include a subpopulation of stars associated with the cluster’s red clump, the $(0.7\text{--}2)M_{\odot}$ (M_{\odot} , solar mass) analogues of the helium-burning horizontal-branch stars (indicated by the orange area). Figure 1b shows the colour–magnitude diagram pertaining to NGC 1696. It also exhibits a bright, young simple-stellar-population sequence characterized by $\log t = 8.70$ (500 Myr), whereas the bulk of the NGC 1696 stars are best represented by an older age of $\log t = 9.18$ (1.5 Gyr). Compared with the cluster’s dominant main sequence, the younger sequence exhibits a colour offset of approximately -0.06 mag (a shift to bluer colours), which is consistent with a stellar population characterized by an enhanced helium abundance of $Y = 0.330$ (33.0% of helium atoms by mass), while $Y = 0.256$ for the bulk of the cluster

stars. A similar result for NGC 411 is shown in Fig. 1c, where we also find an additional, brighter stellar sequence that is well-represented by a younger isochrone of $\log t = 8.50$ (320 Myr) and $Y = 0.400$, compared with $\log t = 9.14$ (1.4 Gyr) and $Y = 0.252$ for the bulk of the cluster’s stellar population. These well-populated, younger and helium-enhanced stellar sequences represent the strongest evidence yet that additional, post-formation starburst events may have occurred in our sample clusters. The enhanced helium abundances may also lead to small changes in the best-fitting cluster ages, but there are currently no appropriate model isochrones available to accurately explore this effect for stellar populations younger than 1 Gyr. Nevertheless, the general sense of helium-enhanced young sequences shown here is robust. To date, no other massive clusters of equivalent age are known to host similarly significant populations of younger stars¹⁰.

We explored whether ‘blue straggler stars’—stars that have been rejuvenated through either stellar collisions or mass transfer in binary stellar systems^{11,12}—could be entirely responsible for the presence of these younger sequences. If they are formed through mass transfer in unresolved, compact binary systems, they would appear brighter and slightly redder than the corresponding isochrone¹³ describing zero-age single stars (stars whose output luminosities are no longer powered by the excess energy gained from gravitational contraction but which are instead driven by nuclear fusion of hydrogen atoms). However, the younger sequences in both clusters are too blue to account for a binary origin and can instead be very well described by single-star isochrones. Given their young ages and the timescales involved in evolution through binary mass transfer, if any of these younger stars are indeed blue stragglers, they will most probably have formed through stellar collisions. Collisionally formed blue stragglers, like secondary stellar generations originating from colliding stellar winds, are expected to be more centrally concentrated than the clusters’ dominant (by number) stellar populations⁶. Figure 2 compares the normalized radial distributions of the young sequences with those of ‘normal’ cluster stars of similar luminosity. The stars in the young sequences are markedly less centrally concentrated than the dominant older population of cluster members.

The more extended nature of the young populations suggests that they may have an external origin. Indeed, since the masses of NGC 1783, NGC 1696 and NGC 411 are only $1.8 \times 10^5 M_{\odot}$, $5.0 \times 10^4 M_{\odot}$ and $3.2 \times 10^4 M_{\odot}$, respectively (refs 14, 15), they are insufficiently massive to efficiently capture the stellar winds from asymptotic-giant-branch stars¹⁶. Note that our NGC 1696 mass is based on extrapolation of the observed stellar luminosity function down to a stellar mass of $0.08 M_{\odot}$ (the minimum mass for hydrogen fusion, somewhat depending on the star’s chemical composition), adopting a Kroupa-like initial mass distribution¹⁷. In addition, the observed upper limits in the mass–age diagrams populated by star clusters in the Magellanic Clouds are well-understood in terms of ‘size-of-sample’ effects. They

¹Kavli Institute for Astronomy and Astrophysics and Department of Astronomy, Peking University, Yi He Yuan Lu 5, Hai Dian District, Beijing 100871, China. ²Key Laboratory for Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China. ³Purple Mountain Observatory, Chinese Academy of Sciences, 2 West Beijing Road, Nanjing 210008, China. ⁴International Space Science Institute – Beijing, 1 Nanertiao, Hai Dian District, Beijing 100190, China. ⁵Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA. ⁶Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA.

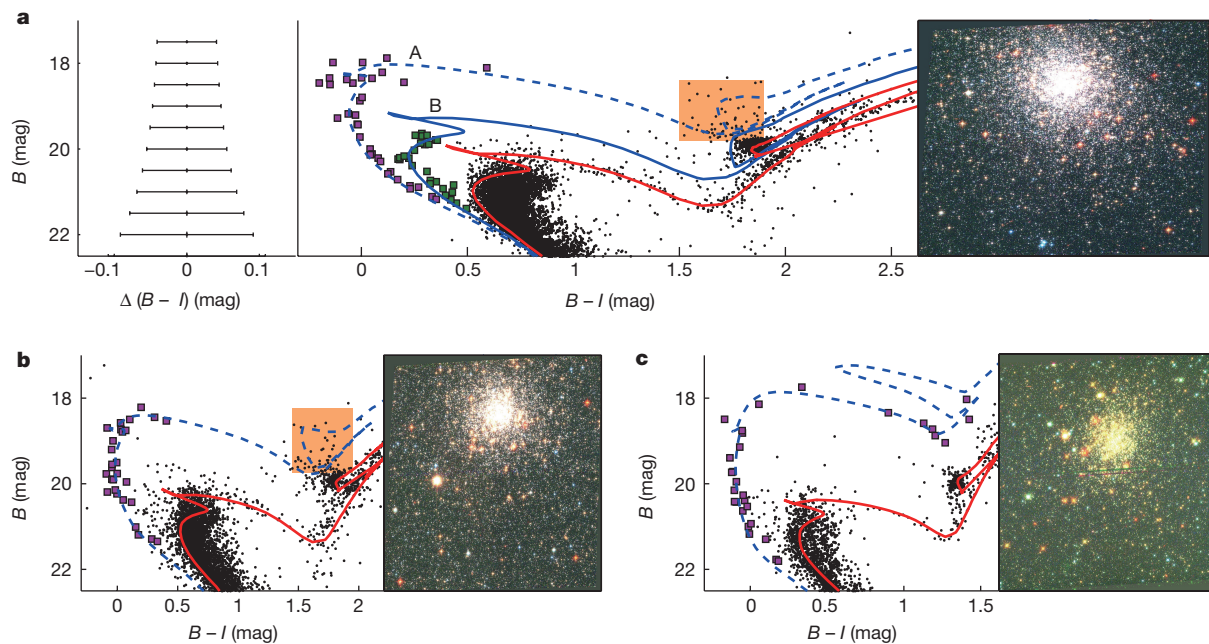


Figure 1 | Colour-magnitude diagrams, including the best-fitting isochrones, and true-colour images for all three clusters. a, NGC 1783. Central panel, purple and dark green squares show sequence A and sequence B stars, respectively; red solid line, blue solid line and blue dashed line show isochrones for $\log t = 9.15$, 8.95 and 8.65, respectively. The orange box indicates the region where red-clump stars associated with sequence A may be found. Left panel, representative $\pm 1\sigma$ measurement uncertainties, note that the x-axis scale is different from that of the main panel. Right panel, true-colour image. **b**, NGC 1696. Left panel, purple

squares, younger sequence; red solid and blue dashed lines, isochrones for $\log t = 9.18$ and 8.70, respectively. Right panel, true colour image. **c**, NGC 411. Left panel, purple squares, younger sequence; red solid and blue dashed lines, isochrones for $\log t = 9.14$ and 8.50, respectively. Right panel, true colour image synthesised from B, V and I band monochromatic images. The measurement uncertainties in **b** and **c** are equivalent to those shown for **a**. For the young sequences in NGC 1696 and NGC 411, enhanced helium abundances have been adopted (see text).

cannot be reconciled with initial cluster populations containing large numbers of young clusters with masses far greater than $10^6 M_\odot$.

Adopting a Kroupa-like initial stellar mass distribution¹⁷, we estimate that the total stellar masses in the younger sequences, down to the hydrogen-burning limit, are $372 M_\odot$ and $250 M_\odot$ (NGC 1783 sequences A and B, respectively), $527 M_\odot$ and $560 M_\odot$ (NGC 1696 and NGC 411, respectively). Compared with the total cluster masses, the young sequences represent mere 0.2%–2.0% mass fractions. The age differences between the NGC 1783 stars in sequences A and B, and those on the main sequence are 440 Myr and 520 Myr, respectively. For NGC 1696 and NGC 411, the age differences between the clusters' young and main sequences are 1.02 Gyr and 1.06 Gyr, respectively.

The Large and Small Magellanic Clouds, the host galaxies of our sample clusters, contain numerous, densely distributed giant molecular-gas clouds^{18,19}. Massive clusters like those targeted here quench their initial star-formation activity on timescales of a few tens of millions of years^{20,21} owing to the occurrence of type II supernovae resulting from the deaths of the most massive first-generation stars^{22,23}. This leaves a cluster embedded in a gas-poor 'cavity'. As the young star cluster moves through the interstellar medium, it could potentially accrete sufficient gas to fuel renewed star formation. In theory, secondary star formation can be triggered^{23–25} and proceed rapidly once the gas density reaches the relevant threshold, for sufficiently low temperatures. This would result in the appearance of a younger 'simple stellar population'. However, to date the reality of this proposed idea has not been confirmed.

Although they are all contained within twice their host clusters' core radii, the observed spatial distributions of all young stellar sequences in our sample clusters are more extended than the cluster's bulk stellar populations. This could indicate that these clusters may have accreted ambient gas, allowing star formation to proceed. We thus explored the expected gas-accretion rate as a function of the local gas density. Indeed, it appears possible for NGC 1783-like clusters to accrete enough gas to form new stars²³ (see Methods).

Almost all Galactic globular clusters host multiple stellar populations. However, it is still unclear whether these latter populations originate from young clusters that formed as single-age stellar populations. Our observations of secondary stellar populations in intermediate-age Magellanic Cloud clusters suggest that the same process giving rise to them may also explain the multiple stellar populations seen in at least some Galactic globular clusters. We aim at addressing this issue in a follow-up study.

Many star clusters in the Magellanic Clouds contain large numbers of stars occupying the colour-luminosity parameter space at bluer colours and brighter luminosities than their main-sequence turn-off regions^{26,27}. These clusters include NGC 121, NGC 1652, NGC 1751, NGC 1795, NGC 1806, NGC 1846, NGC 1852, NGC 1917, NGC 1978, NGC 2121, NGC 2154, NGC 339, NGC 416, NGC 419, Hodge 7, Kron 3, Lindsay 1 and Lindsay 38. These brighter and bluer stars are usually dismissed as residual field-star contamination. However, if well-populated younger sequences are, in fact, embedded in this parameter space, the presence of such simple stellar populations indicates that many clusters may have experienced starburst events some time after their initial formation epoch. Among the clusters highlighted here, some—including NGC 1806, NGC 1846, Lindsay 38 and NGC 419—appear to exhibit features similar to NGC 1783, NGC 1696 and NGC 411, although at a lower level of significance. Particularly for Lindsay 38 and NGC 419, the radial distributions of the bright stars beyond their main-sequence turn-off regions are known to be less concentrated than the bulk cluster stars with similar luminosities²⁸, and hence a population of blue stragglers probably cannot explain the properties of all those stars. Some may have originated from gas accretion. These clusters will be targeted in our follow-up studies. Our discovery of clear, well-populated young sequences in NGC 1783, NGC 1696 and NGC 411 has revealed that star clusters may indeed have the capacity to accrete gas from their environment. This could, in fact, be the most important route to form secondary stellar populations in young massive clusters.

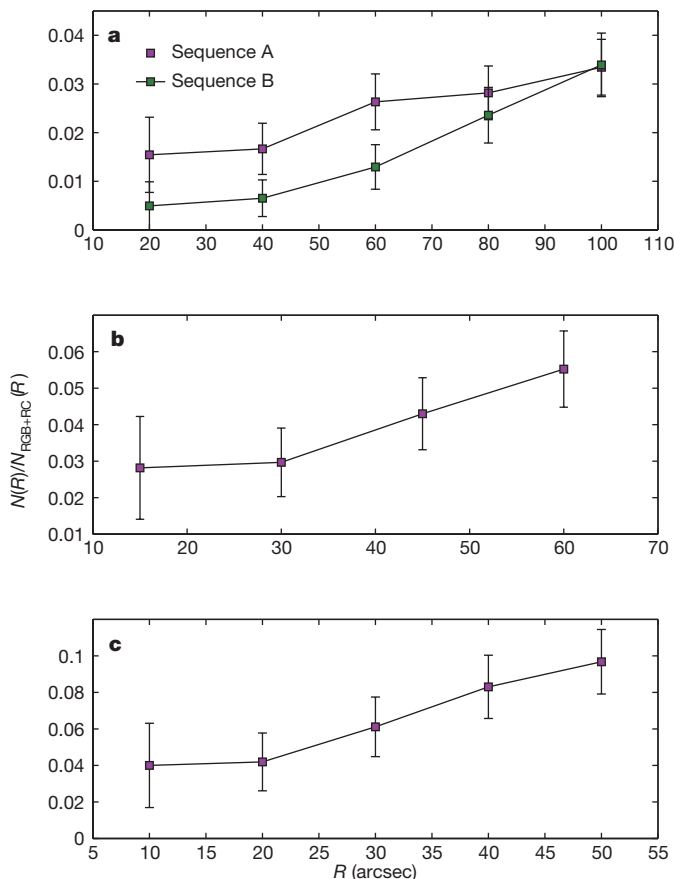


Figure 2 | Normalized radial distributions of the young sequences with respect to normal cluster stars of similar luminosity. $N(R)$ represents the number of stars in the young sequences within radius R , whereas $N_{\text{RGB+RC}}(R)$ represents the total number of red-giant branch and red clump stars within R . RGB, red-giant branch; RC, red clump. **a**, NGC 1783; **b**, NGC 1696; and **c**, NGC 411. The error bars reflect Poissonian $\pm 1\sigma$ uncertainties.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 August; accepted 19 November 2015.

- Gratton, R. G., Carretta, E. & Bragaglia, A. Multiple populations in globular clusters. Lessons learned from the Milky Way globular clusters. *Astron. Astrophys. Rev.* **20**, 50 (2012).
- Piotto, G. *et al.* A triple main sequence in the globular cluster NGC 2808. *Astron. Astrophys. J.* **661**, L53–L56 (2007).
- Milone, A. P. *et al.* Multiple stellar populations in 47 Tucanae. *Astron. Astrophys. J.* **744**, 58–79 (2012).
- Bedin, L. R. *et al.* ω Centauri: the population puzzle goes deeper. *Astron. Astrophys. J.* **605**, L125–L128 (2004).
- D'Ercole, A., Vesperini, E., D'Antona, F., McMillan, S. L. W. & Recchi, S. Formation and dynamical evolution of multiple stellar generations in globular clusters. *Mon. Not. R. Astron. Soc.* **391**, 825–843 (2008).
- Bekki, K. Secondary star formation within massive star clusters: origin of multiple stellar populations in globular clusters. *Mon. Not. R. Astron. Soc.* **412**, 2241–2259 (2011).
- Renzini, A. Origin of multiple stellar populations in globular clusters and their helium enrichment. *Mon. Not. R. Astron. Soc.* **391**, 354–362 (2008).

- Bastian, N. & Lardo, C. Globular cluster mass loss in the context of multiple populations. *Mon. Not. R. Astron. Soc.* **453**, 357–364 (2015).
- Marigo, P. *et al.* Evolution of asymptotic giant branch stars. II. Optical to far-infrared isochrones with improved TP-AGB models. *Astron. Astrophys.* **482**, 883–905 (2008).
- Bastian, N., Cabrera-Ziri, I., Davies, B. & Larsen, S. S. Constraining globular cluster formation through studies of young massive clusters. I. A lack of ongoing star formation within young clusters. *Mon. Not. R. Astron. Soc.* **436**, 2852–2863 (2013).
- Lu, P., Deng, L. & Zhang, X. Blue straggler formation via close binary mass transfer. *Mon. Not. R. Astron. Soc.* **409**, 1013–1021 (2010).
- Hills, J. G. & Day, C. A. Stellar collisions in globular clusters. *Astrophys. J.* **17**, L87–L93 (1976).
- Li, C., de Grijs, R. & Deng, L. The binary fractions in the massive young Large Magellanic Cloud star clusters NGC 1805 and NGC 1818. *Mon. Not. R. Astron. Soc.* **436**, 1497–1512 (2013).
- Hunter, D. A., Elmegreen, B. G., Dupuy, T. J. & Mortonson, M. Cluster mass function in the Large and Small Magellanic Clouds: fading and size-of-sample effects. *Astron. J.* **126**, 1836–1848 (2003).
- McLaughlin, D. E. & van der Marel, R. P. Resolved massive star clusters in the Milky Way and its satellites: brightness profiles and a catalog of fundamental parameters. *Astrophys. J.* **161** (Suppl.), 304–360 (2005).
- Recchi, S. & Danziger, I. J. Self-enrichment in globular clusters. I. An analytic approach. *Astron. Astrophys.* **436**, 145–154 (2005).
- Kroupa, P. On the variation of the initial mass function. *Mon. Not. R. Astron. Soc.* **322**, 231–246 (2001).
- Fukui, Y. & Kawamura, A. Molecular clouds in nearby galaxies. *Annu. Rev. Astron. Astrophys.* **48**, 547–580 (2010).
- Hopkins, P. F., Quataert, E. & Murray, N. The structure of the interstellar medium of star-forming galaxies. *Mon. Not. R. Astron. Soc.* **421**, 3488–3521 (2012).
- de Grijs, R. & Parmentier, G. The long-term survival chances of young massive star clusters. *Chin. J. Astron. Astrophys.* **7**, 155–186 (2007).
- de Grijs, R. A revolution in star cluster research: setting the scene. *Phil. Trans. R. Soc. A* **368**, 693–711 (2010).
- Shustov, B. M. & Wiebe, D. S. First supernovae in Galactic globular clusters. *Mon. Not. R. Astron. Soc.* **319**, 1047–1055 (2000).
- Conroy, C. & Spergel, D. N. On the formation of multiple stellar populations in globular clusters. *Astrophys. J.* **726**, 36–48 (2011).
- Pflamm-Altenburg, J. & Kroupa, P. Recurrent gas accretion by massive star clusters, multiple stellar populations and mass thresholds for spheroidal stellar systems. *Mon. Not. R. Astron. Soc.* **397**, 488–494 (2009).
- Naiman, J. P., Ramirez-Ruiz, E. & Lin, D. N. C. External mass accumulation onto core potentials: implications for star clusters, galaxies, and galaxy clusters. *Astrophys. J.* **735**, 25–37 (2011).
- Milone, A. P., Bedin, L. R., Piotto, G. & Anderson, J. Multiple stellar populations in Magellanic Cloud clusters. I. An ordinary feature for intermediate age globulars in the LMC? *Astron. Astrophys.* **497**, 755–771 (2009).
- Rich, R. M., Shara, M. M. & Zurek, D. New photometry for the intermediate-age Large Magellanic Cloud globular cluster NGC 2121 and the nature of the LMC age gap. *Astron. J.* **122**, 842–848 (2001).
- Glatt, K. *et al.* Age determination of six intermediate-age Small Magellanic Cloud star clusters with HST/ACS. *Astron. J.* **136**, 1703–1727 (2008).

Acknowledgements We thank F. Ferraro, P. Kroupa and X. K. Liu for discussions. Partial financial support for this work was provided by the National Natural Science Foundation of China through grants 11073001, 11373010 and 11473037. C.L. was also partially supported by the Chinese Academy of Sciences (grant XDB09000000) and the 973 Program (grant 2014CBB45700). A.M.G. was funded by a National Science Foundation Astronomy and Astrophysics Postdoctoral Fellowship under award no. AST-1302765. C.-A.F.-G. was supported by a National Science Foundation grant AST-1412836.

Author Contributions C.L., R.d.G. and L.D. jointly designed and coordinated this study. C.L. performed the data reduction. C.L., R.d.G., Y.X. and Y.H. collaborated on the detailed analysis. L.D. provided ideas that improved the study's robustness. A.M.G. and C.-A.F.-G. led the theoretical analysis of the gas-accretion physics. All authors read, commented on and jointly approved submission of this article.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.L. (joshuali@pku.edu.cn or licy@pmo.ac.cn).

METHODS

Observational data. NGC 1783. The NGC 1783 observations were obtained as part of Hubble Space Telescope (HST) general-observer programme GO-10595 (principal investigator, P. Goudfrooij), using the Advanced Camera for Surveys/Wide Field Channel (ACS/WFC). The cluster was observed through the F435W and F814W filters (with central wavelengths of 435 nm and 814 nm, respectively), which correspond approximately to the Johnson–Cousins B and I bands, respectively, and which will be referred to as such henceforth. Short-exposure images were observed for 90 s in the B band and 8 s in the I band, while long-exposure images were observed for 680 s in both bands.

Because NGC 1783 has an extended core (see below), the cluster region occupies almost the entire image. Therefore, we obtained an additional set of observations towards the southeast of NGC 1783 as a representative field region (HST programme GO-12257; principal investigator, L. Girardi). Its centre is located at a distance of more than 300 arcsec from the cluster centre, so that it is unlikely to be significantly affected by tidally stripped cluster stars. The data sets pertaining to the field were also obtained with the ACS/WFC in the F435W and F814W filters. Their total exposure times are 700 s and 720 s for the B and I bands, respectively. These exposure times are sufficient to resolve the young sequences.

NGC 1696. The NGC 1696 data sets were also obtained as part of HST programme GO-10595, using the ACS/WFC. The cluster was observed through the F435W and F814W filters. The images of NGC 1696 are also composed of short- and long-exposure frames. The long (short) exposure times in the B and I bands were 680 s (90 s) and 680 s (8 s), respectively. Since NGC 1696 is not as extended as NGC 1783 (see below), a representative field region was selected close to the edge of the NGC 1696 images. Few studies to date have targeted NGC 1696. Therefore, we derived the physical parameters of NGC 1696 ourselves.

NGC 411. The data sets of NGC 411 were obtained as part of HST programme GO-12257, using the Wide Field Camera-3 (WFC3). The cluster was observed through the F475W and F814W filters. The F475W filter is centred at a wavelength of 475 nm; its transmission curve also corresponds approximately to that of the Johnson B band. The exposure times in both the B and I bands were 700 s. NGC 411 has a relatively small core (see below), which allowed us to select a representative field region close to the edge of our science images.

Photometry and data reduction. We used two independent software packages to perform point-spread-function (PSF) photometry, including DAOPHOT²⁹ within the IRAF environment and DOLPHOT^{30–32}. Our stellar catalogues are based on the DAOPHOT results. We performed the DOLPHOT analysis for comparison, to ensure that our final photometry is not biased.

The DAOPHOT-generated raw stellar catalogue contains a sharpness parameter, which describes the goodness of the PSF fit²⁹. For a ‘good’ star, the sharpness should be close to 0. We thus constrained our sample to stars with a sharpness between -0.5 and 0.5 , which removed approximately 4% of objects from our catalogue. We carefully checked the resulting colour–magnitude diagram and found that the main features of interest were not affected by this selection. For NGC 1783 and NGC 1696, we merged the stellar samples resulting from the short and long exposure times, carefully cross-referencing both catalogues to avoid duplication of objects in the combined output catalogue. For NGC 1783, the short-exposure-time catalogue contributes very little to sequences A and B. For NGC 1696, the short-exposure-time stars contribute only marginally to the feature of interest.

Determination of the cluster and field regions. We divided the stellar spatial distribution into 15–20 bins along both the right ascension (α_{J2000}) and declination (δ_{J2000}) axes. We varied the bin numbers to ensure that statistical scatter would not significantly affect the shape of the number-density distributions. We used a Gaussian function to fit the latter along both axes and defined the closest positional coincidence of both Gaussian peaks as the cluster centre. The NGC 1783 cluster centre is located at $\alpha_{J2000} = 04^{\text{h}} 59^{\text{m}} 08.47^{\text{s}}$, $\delta_{J2000} = -65^{\circ} 59' 17.81''$. For NGC 1696, the centre coordinates are $\alpha_{J2000} = 05^{\text{h}} 02^{\text{m}} 11.16^{\text{s}}$, $\delta_{J2000} = -67^{\circ} 59' 07.66''$, while for NGC 411, $\alpha_{J2000} = 01^{\text{h}} 07^{\text{m}} 56.22^{\text{s}}$, $\delta_{J2000} = -71^{\circ} 46' 04.40''$. Our NGC 1783 and NGC 411 cluster centres are very close to the cluster centres determined previously^{33,34}. We used a Monte Carlo method to examine the spatial distribution of the NGC 1696 stars and estimate the areas of annuli at different radii. The stellar number density in each annulus is $N(R)/A(R)$, where $N(R)$ is the number of stars observed in an annulus with radius R , and $A(R)$ is the area of the annulus. We defined the clusters’ (2D-projected) core radii as those radii where the density profiles drop to half the respective central densities. We selected the areas contained within 2 core radii as the cluster regions. NGC 1783 has a large core radius (45–50 arcsec, which is identical to that adopted by ref. 35, although our core radius is slightly larger than their value of 36.7 arcsec. We compared our catalogue with theirs and found that our photometry is deeper and, hence, contains more faint stars). The core radius of NGC 1696 is smaller (~ 30 arcsec), while the core radius of NGC 411 is approximately half that of NGC 1783 (20–25 arcsec).

Hence, the cluster radii we adopted for NGC 1783, NGC 1696 and NGC 411 were 100 arcsec, 60 arcsec and 50 arcsec, respectively. We selected these radii as cluster radii, because of the need to avoid background contamination as much as possible, while simultaneously ensuring statistically robust results. In Extended Data Fig. 1, we present the stellar number-density profiles.

For NGC 1783, we synthesized the field region’s colour–magnitude diagram based on a combination of our observations of the cluster’s periphery on the image containing the cluster (‘field 1’) and the field region towards the southeast (‘field 2’). Because the features of interest are very bright, we only consider stars with $B < 23$ mag. For these stars, NGC 1783 is compact (its core radius is 25 arcsec) and field 1 indeed adequately represents the background. We found that the eastern part of field 2 exhibits a clear, slightly brighter stellar sequence parallel to the main sequence, indicating a significant population of unresolved binary systems. Since the field’s stellar number density is low, blending of unrelated stars along the line of sight is negligible; instead, this binary population may reflect contamination by a nearby star-forming region. We hence selected the eastern part of field 2 as a representative field region. We selected three rectangles, each covering an area of 1,000 pixels \times 1,000 pixels (~ 50 arcsec \times 50 arcsec), near the edge of field 1, as well as four rectangles from the western part of field 2 (covering the same area), to construct a complete, combined field region. To assign equal weights to the stellar catalogues from both regions, we randomly selected three-quarters of the full sample of stars detected in the four rectangles of field 2’s stellar catalogue. The combined stellar catalogue based on these seven rectangles represents the synthesized field region’s colour–magnitude diagram. The cluster region is roughly 1.6 times larger than the field region (see the left-hand panel of Extended Data Fig. 2).

For NGC 1696, we selected an area of 600 pixels \times 4,000 pixels (~ 30 arcsec \times 200 arcsec) near the edge of the image as a representative field region. The NGC 1696 cluster region is roughly 1.7 times larger than the field region. The selected cluster and field regions pertaining to NGC 1696 are shown in the middle panel of Extended Data Fig. 2.

For NGC 411, we selected a field region covering an area of 800 pixels \times 3,500 pixels (~ 32 arcsec \times 140 arcsec) from the cluster’s periphery. We avoided regions that were located close to the edge of the image, where the photometric quality is markedly inferior, probably owing to the relatively large offsets between the individual science images, combined with instrumental propagation effects. Although our selection may include some tidally stripped cluster stars, this does not affect the magnitude range of interest, which is bright. The NGC 411 cluster region is roughly 1.8 times larger than the field region. The selected cluster and field regions of NGC 411 are shown in the right-hand panel of Extended Data Fig. 2.

Reducing background contamination and isochrone fitting. Once we had obtained the cluster and background colour–magnitude diagrams, we generated a common magnitude \times colour grid with cell sizes of 0.5 mag \times 0.25 mag, spanning the ranges from $(B - I) = -2.5$ mag to 3.5 mag and from $B = 16$ mag to 27 mag. This range is sufficiently large to cover the full colour–magnitude diagrams of our target clusters. The cell size is relatively large for the main sequences, because it was specifically designed to be practically useful in the regions occupied by the young, blue sequences, where the stellar number density is lower. We counted the number of background stars in each cell and calculated the number of possible contaminating stars in the same cell (corrected for differences in areas covered), which we then randomly removed. We confirmed that varying the cell sizes from 0.3 mag \times 0.15 mag to 0.5 mag \times 0.25 mag for NGC 1783 would not affect the significance of any of the features of interest. For NGC 1696, the typical practically useful cell sizes range from 0.4 mag \times 0.4 mag to 0.5 mag \times 0.5 mag, while for NGC 411, viable cell sizes range from roughly 0.3 mag \times 0.15 mag to 0.5 mag \times 0.25 mag. Adopting much larger or smaller cell sizes would erase the observed sequences. We carefully examined the performance of our decontamination method and found that the observed features do not depend on the cluster or field regions selected: see Extended Data Figs 3 and 4, where we use NGC 1783 as benchmark. In Extended Data Fig. 3, we show the field-decontaminated colour–magnitude diagrams pertaining to three different samples of cluster members, at radii $R \geq 30''$, $R \geq 60''$ and $R \geq 90''$. They all exhibit distinct younger sequences. In Extended Data Fig. 4, we select representative field regions from different images (two from a separate image and one from the image which also contains the cluster itself). The observed young sequences remain clearly visible for all three background regions adopted. In Extended Data Fig. 5, we show the decontaminated colour–magnitude diagrams resulting from adoption of different grid sizes. This figure shows that the observed features are almost independent of grid size. Extensive tests also showed that the sequences found in the NGC 1696 and NGC 419 colour–magnitude diagrams are similarly well-defined. This confirms that the observed sequences are physically real rather than caused by statistical sampling effects. We also found that, for the adopted cell sizes, the reduced colour–magnitude distributions are similar in appearance to the real background colour–magnitude diagrams. Our adopted

method therefore performs adequately. From Extended Data Fig. 6a, e and i, one can deduce that these bright sequences are indeed already embedded in the unreduced colour–magnitude diagrams.

We obtained best fits to all observed sequences, including the clusters' main sequences, based on matching the observations with theoretical stellar isochrones⁹. Similarly to other intermediate-age star clusters, NGC 1783, NGC 1696 and NGC 411 display extended main-sequence turn-off regions^{26,36}, which renders determination of their main-sequence ages difficult. However, it has been reported that ages of such intermediate-age star clusters can be constrained by consideration of their tight subgiant branches^{37,38} (but see ref. 39 for an opposing view), which represent the stellar evolutionary stage stars enter once they have exhausted the hydrogen in their cores through nuclear fusion. Indeed, all of our sample clusters exhibit tight subgiant branches. The final age determination yields $\log t = 9.15$, 9.17 and 9.14 for NGC 1783, NGC 1696 and NGC 411, respectively.

We next determined the ages of each of the blue sequences. For NGC 1783, sequences A and B are characterized by ages of $\log t = 8.65$ and 8.95, respectively. The NGC 1783 main-sequence stars, as well as those in sequences A and B, share the same metal abundance, $Z = 0.008$ (40% of solar metallicity), and visual extinction, $A_V = 0.06$ mag. We adopted a true distance modulus for NGC 1783 of $(m - M)_0 = 18.46$ mag (corresponding to a distance of 49.2 kpc). The young sequence in NGC 1696 is adequately characterized by a $\log t = 8.70$ isochrone with the same metallicity as the cluster's main sequence, $Z = 0.004$ (20% solar). However, the young sequence appears 0.06 mag bluer than the zero-age main sequence of the cluster's bulk stellar population. An enhanced helium abundance ($Y = 0.330$) can explain the colour offset⁴⁰, where $Y = 0.256$ for the cluster's main sequence. The adopted extinction and distance modulus for NGC 1696 are $A_V = 0.10$ mag and $(m - M)_0 = 18.50$ mag (50.1 kpc), respectively. The young sequence in NGC 411 has an age of $\log t = 8.50$. It has the same metallicity as the cluster's main sequence, $Z = 0.002$. Its young sequence is also very blue, which can again only be explained if it is characterized by an enhanced helium abundance of $Y = 0.400$ compared with $Y = 0.252$ for the cluster's main sequence. Foreground extinction of $A_V = 0.25$ mag is appropriate and our adopted true distance modulus is $(m - M)_0 = 18.90$ mag (60.3 kpc).

Blue straggler stars as possible origins of the younger sequences. One possible explanation for these bright sequences is that they are composed of blue straggler stars. Therefore, we investigated their relative radial concentration with respect to stars that have similar luminosities. In NGC 1783, NGC 1696 and NGC 411, the latter stars are mostly red-giant-branch and red-clump stars: see Extended Data Fig. 7.

However, as shown in Fig. 2, the stars defining the bright sequences are all less centrally concentrated than red-giant stars with similar luminosities. If they are genuine blue stragglers, irrespective of their origin, they are expected to be more centrally concentrated than similar-luminosity red-giant stars, because blue stragglers are expected to be more massive than red giants. Dynamical interactions are also unlikely to have redistributed all blue stragglers to the outskirts of our sample clusters, since the typical dynamical timescales⁴¹ are much longer than the clusters' current ages. In addition, the flat cluster cores observed for NGC 1783, NGC 1696 and NGC 411 render the probability of core-collapse events having occurred very unlikely. Core collapse would produce a cuspy radial density profile⁴² and such a process is a prerequisite for the presence of a coeval collisional blue straggler population. The more extended radial distributions of the stars in the younger sequences compared with the dominant cluster populations argue against stellar collisions in the cluster cores having played an important role.

Gas accretion from the interstellar medium. We followed the method of ref. 23 to estimate the regions of parameter space where an NGC 1783-like star cluster with a mass of $1.8 \times 10^5 M_\odot$ and a half-mass radius of 11.4 pc (ref. 43) could accrete the required mass to form two additional generations of stars: one containing $250 M_\odot$ in stars some 520 Myr after the initial star-formation event, and a subsequent generation composed of $370 M_\odot$ of stars 440 Myr later. The key equations in ref. 23 are equations (3), (5) and (8). For all calculations we assumed a star-formation efficiency of 10% (which implies that only 10% of the available gas is converted into stars). We explored two processes through which a cluster can accrete gas from

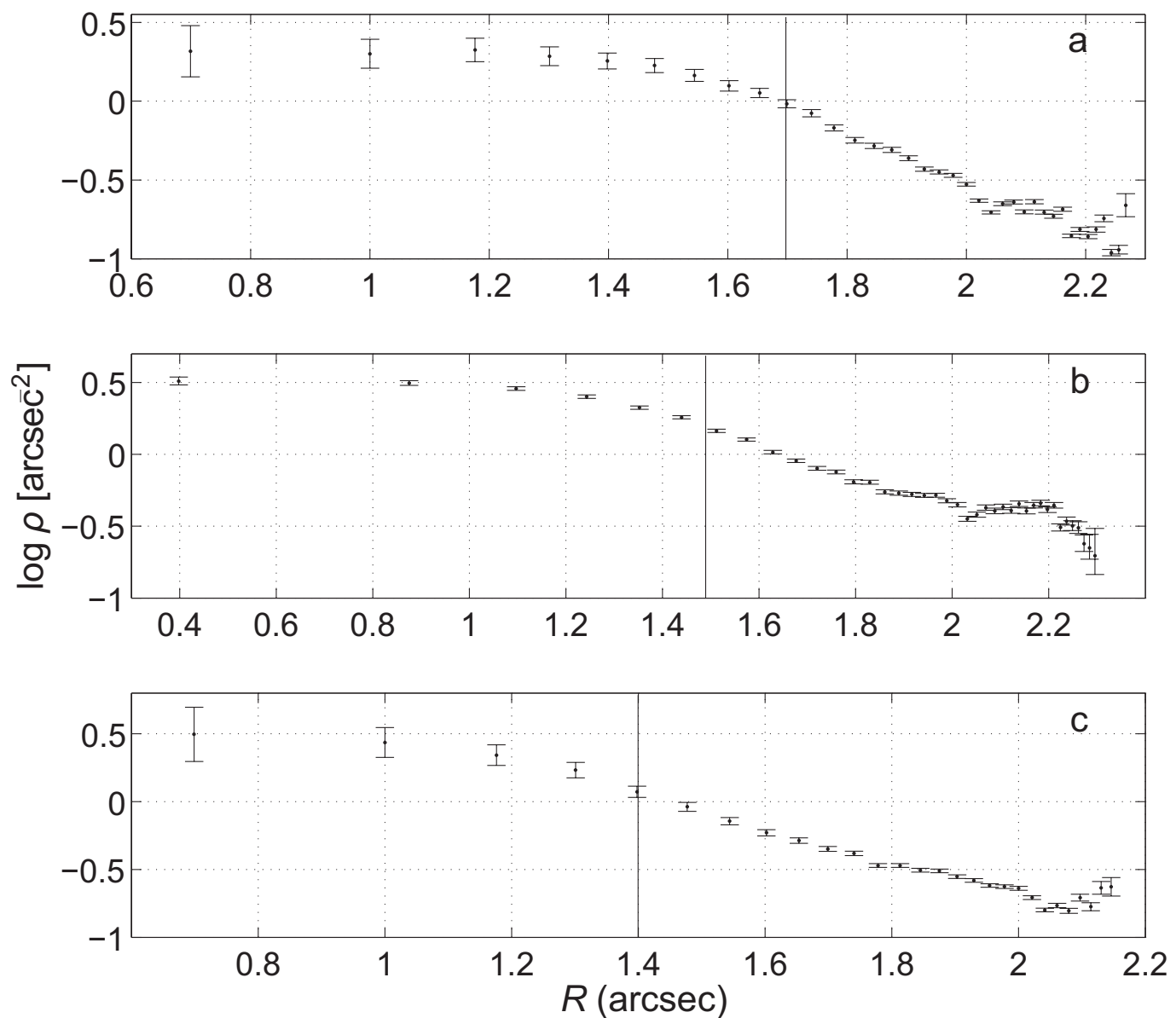
the interstellar medium²³, one due to gravity ('Bondi accretion'), and one due to sweeping up material as the cluster orbits around its host galaxy's centre. The latter process involves some initial intracluster gas which interacts collisionally with the interstellar gas. The accretion rates from both of these processes depend on both the gas density, n , and the relative velocity of cluster and gas, V . Ram-pressure stripping can limit the accretion, and this process also depends on n and V .

Extended Data Fig. 8 shows the allowed (shaded) regions of parameter space in (n, V) that would enable a cluster to accrete the desired amount of gas in the period between star-formation events. This figure assumes accretion from a volume-filling interstellar medium over the entire duration between respective generations (Bondi accretion). The Bondi line is curved, because this accretion rate also depends on the gas sound speed, which we have assumed to be 10 km s^{-1} (that is, gas at a temperature of $\sim 10^4 \text{ K}$). Anything to the upper right of the 'Ram' line will strip the gas from the cluster. Below that line, and for a given density, the Bondi line defines the upper limit to the cluster's bulk velocity that would be allowed for the cluster to accrete this amount of gas. In other words at a velocity below the curve, the cluster would accrete gas more quickly, and, for instance, if the star-formation efficiency were less than 10%, it could still form the same mass of stars over the same period. Conversely, at a given density, sweeping up of gas is more efficient at larger velocities, so the dashed line shows the lower limit to the velocity that would be required to reach the desired amount of gas.

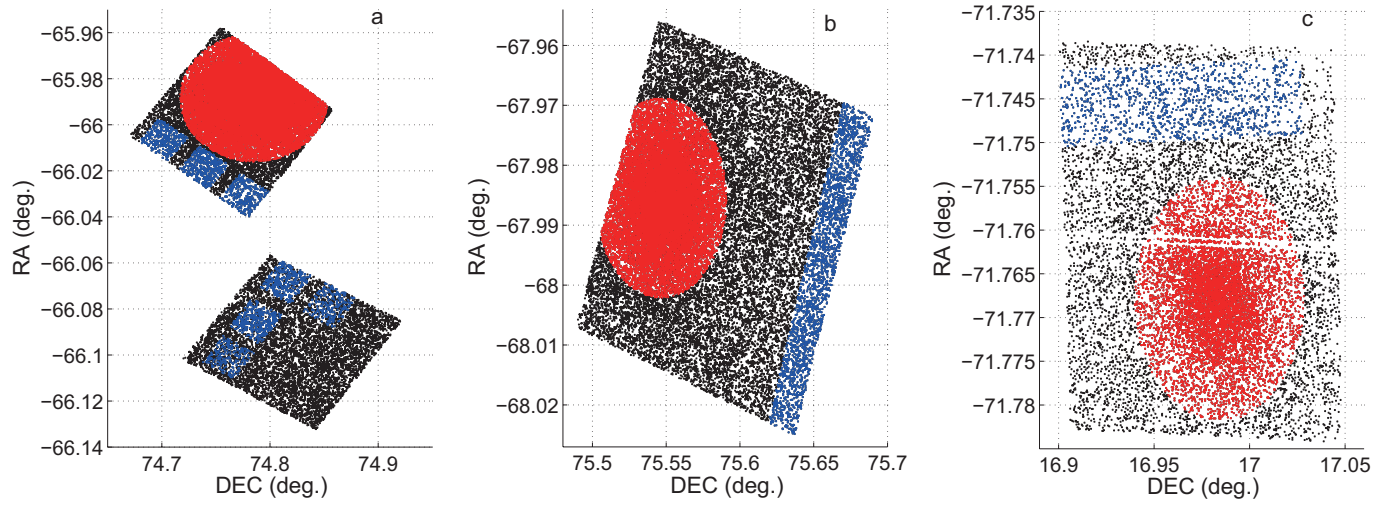
Extended Data Fig. 9 shows two examples of how the accreted mass could accumulate over time for the Bondi regime (solid lines) and the sweeping regime (dashed lines). For the Bondi regime, we used a relative velocity of 4 km s^{-1} and a density of 0.3 cm^{-3} . For the sweeping regime, we used a velocity of 50 km s^{-1} and a density of 0.05 cm^{-3} (which is similar to the rotation velocity at the position of NGC 1783 with respect to the Large Magellanic Cloud's centre⁴⁴). The data points show these masses, with $180 M_\odot$ uncertainties divided by the star-formation efficiency of 10%.

Sample size. No statistical methods were used to predetermine sample size.

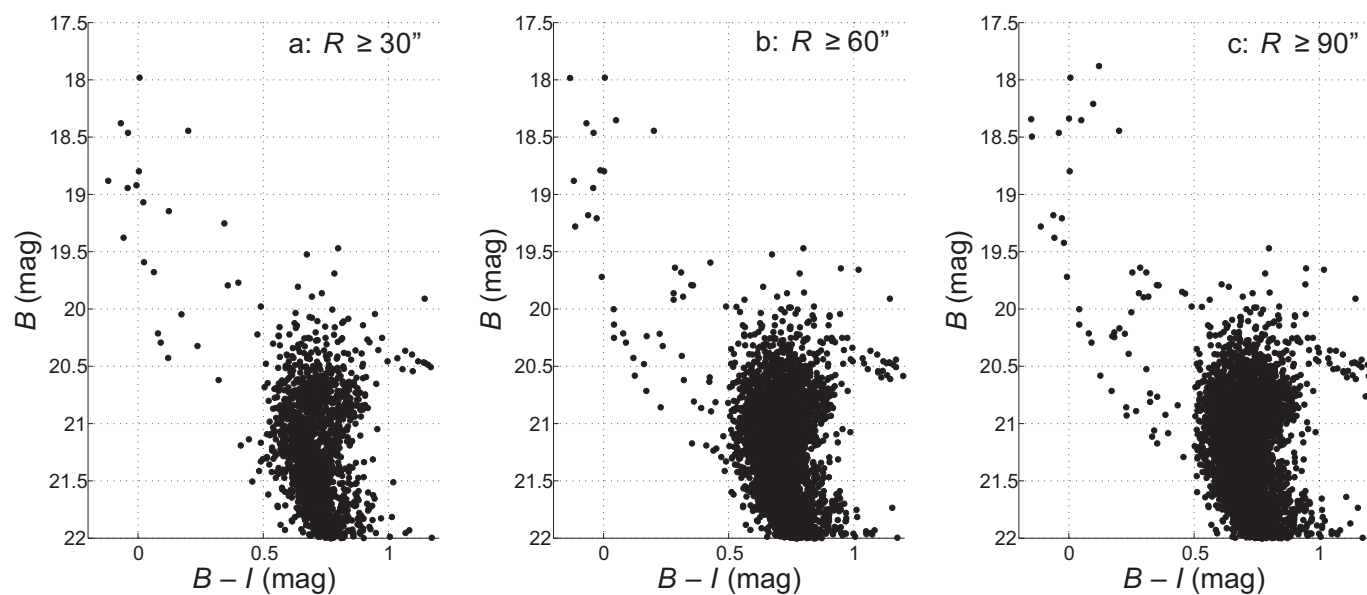
29. Davis, L. E. A reference guide to the IRAF/DAOPHOT package. <http://iraf.noao.edu/ftp/docs/daoreman.ps.Z> (1994).
30. Dolphin, A. DOLPHOT user's guide, version 2.0. <http://americano.dolphinim.com/dolphot/dolphot.pdf> (2013).
31. Dolphin, A. DOLPHOT/ACS user's guide, version 2.0. <http://americano.dolphinim.com/dolphot/dolphotACS.pdf> (2011).
32. Dolphin, A. DOLPHOT/WFC3 user's guide, version 2.0. <http://americano.dolphinim.com/dolphot/dolphotWFC3.pdf> (2011).
33. Werchan, F. & Zaritsky, D. The star clusters of the Large Magellanic Cloud: structural parameters. *Astron. J.* **142**, 48–58 (2011).
34. Rafelski, M. & Zaritsky, D. The star clusters of the Small Magellanic Cloud: age distribution. *Astron. J.* **129**, 2701–2713 (2005).
35. Niederhofer, F. *et al.* Controversial age spreads from the main sequence turn-off and red clump in intermediate-age clusters in the LMC. *Astron. Astrophys.* (in the press); preprint at <http://arXiv.org/abs/1510.08476> (2015).
36. Girardi, L. *et al.* An extended main-sequence turn-off in the Small Magellanic Cloud star cluster NGC 411. *Mon. Not. R. Astron. Soc.* **431**, 3501–3509 (2013).
37. Li, C., de Grijs, R. & Deng, L. The exclusion of a significant range of ages in a massive star cluster. *Nature* **516**, 367–369 (2014).
38. Bastian, N. & Niederhofer, F. The morphology of the subgiant branch and red clump reveal no sign of age spreads in intermediate-age clusters. *Mon. Not. R. Astron. Soc.* **448**, 1863–1873 (2015).
39. Goudfrooij, P. *et al.* On the interpretation of subgiant branch morphologies of intermediate-age star clusters with extended main sequence turnoffs. *Mon. Not. R. Astron. Soc.* **450**, 1693–1704 (2015).
40. Dotter, A. *et al.* The Dartmouth stellar evolutionary database. *Astrophys. J.* **178** (Suppl.), 89–101 (2008).
41. Binney, J. & Tremaine, S. *Galactic Dynamics* (Princeton Univ. Press, 1998).
42. Ashman, K. M. & Zepf, S. E. *Globular Cluster Systems* (Cambridge Astrophys. Ser. Vol. 30, Cambridge Univ. Press, 1998).
43. Goudfrooij, P., Girardi, L. & Kozhurina-Platais, V. Extended main sequence turnoffs in intermediate-age star clusters: a correlation between turnoff width and early escape velocity. *Astrophys. J.* **797**, 35–56 (2014).
44. van der Marel, R. P., Alves, D. R., Hardy, E. & Suntzeff, N. B. New understanding of Large Magellanic Cloud structure, dynamics, and orbit from carbon star kinematics. *Astron. J.* **124**, 2639–2663 (2002).



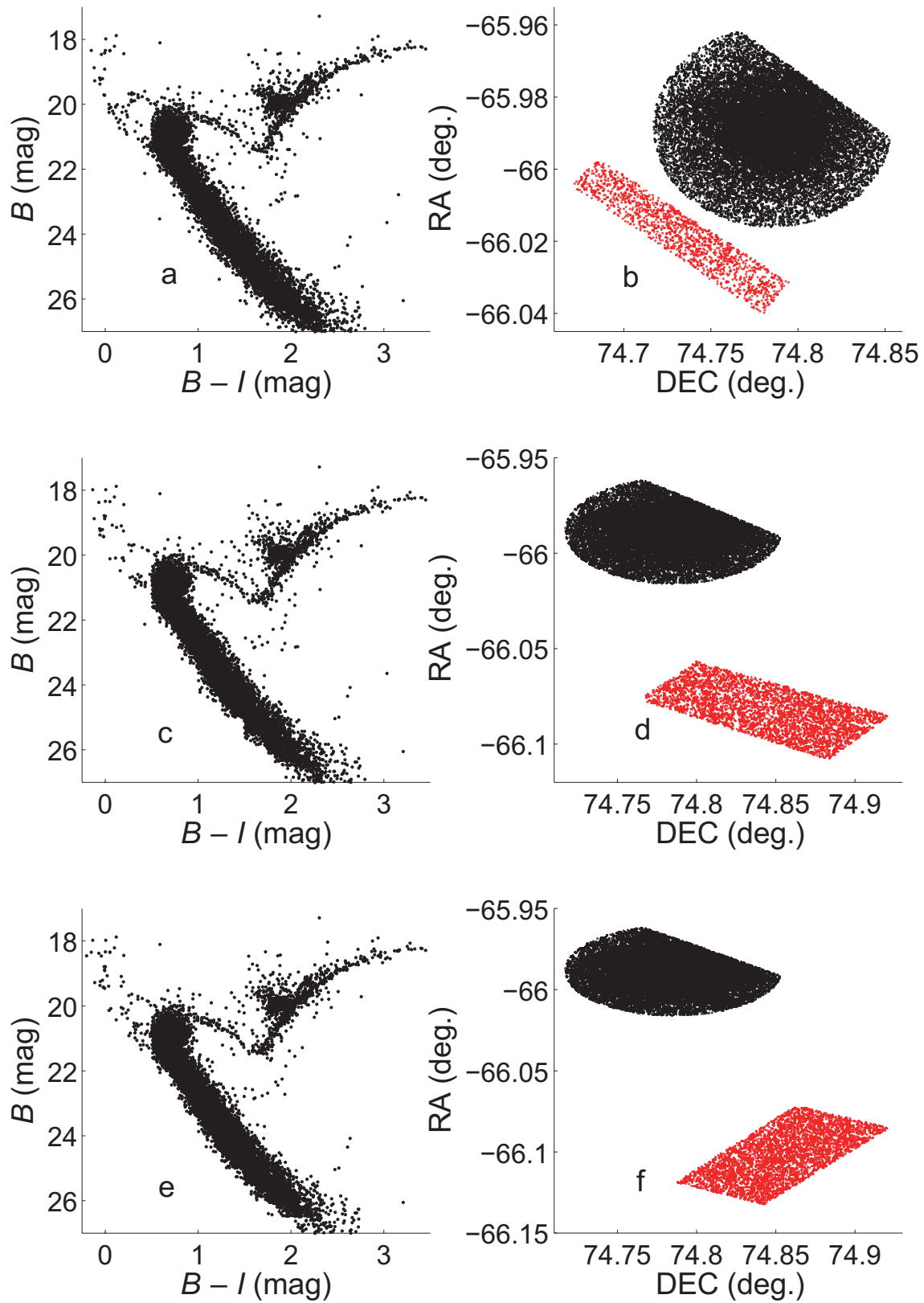
Extended Data Figure 1 | Stellar number-density profiles. Here ρ is the number density of stars at a given radius R . **a**, NGC 1783. The vertical solid line indicates the cluster's core radius, that is, the position where the number density decreases to half the central value. The $\pm 1\sigma$ uncertainties shown are due to Poisson noise. **b**, **c**, As **a**, but for NGC 1696 (**b**) and NGC 411 (**c**).



Extended Data Figure 2 | Spatial distributions of cluster and field stars. Red and blue points represent cluster stars and the adopted field stars, respectively. The black dots are all observed stars except for the cluster and field stars. RA, right ascension; dec., declination. **a**, NGC 1783; **b**, NGC 1696; and **c**, NGC 411.



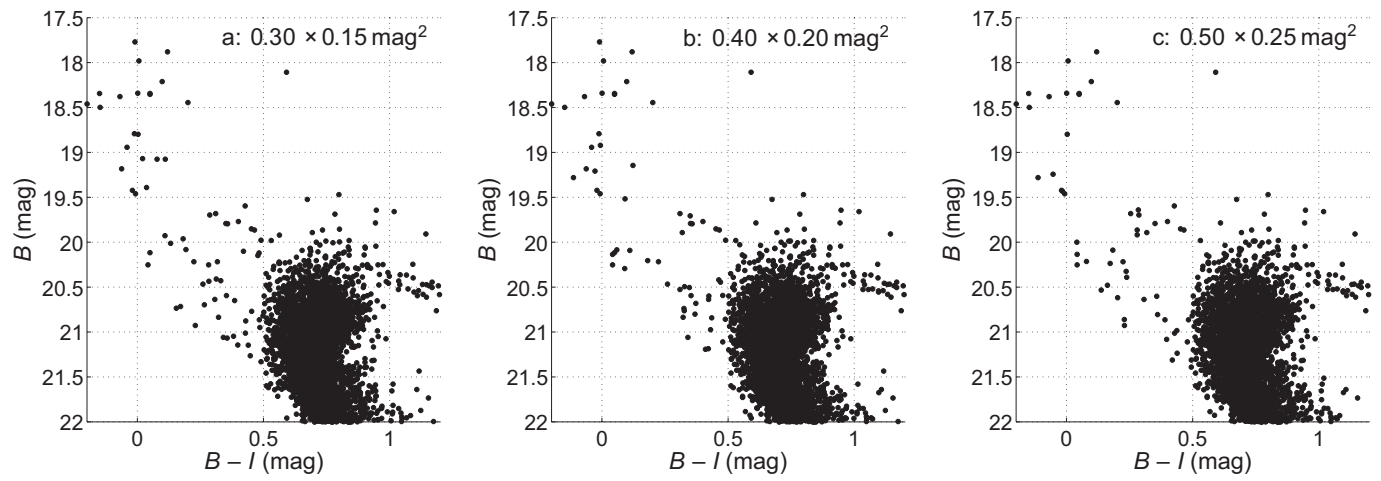
Extended Data Figure 3 | Field-star-decontaminated colour-magnitude diagrams for samples of stars at different radii in NGC 1783. a, $R \geq 30''$; b, $R \geq 60''$; and c, $R \geq 90''$.



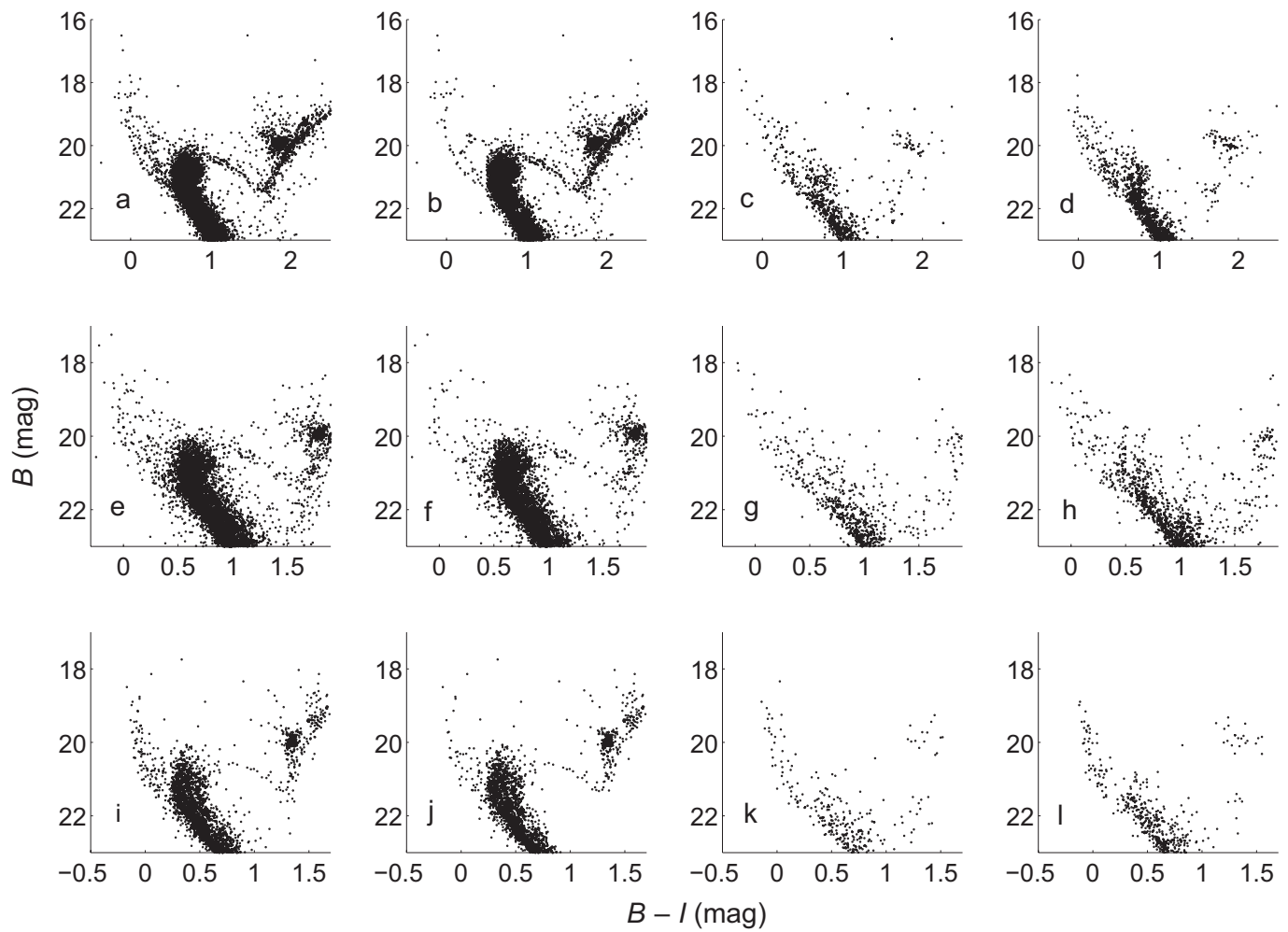
Extended Data Figure 4 | Field-star-decontaminated colour-magnitude diagrams of NGC 1783 for three different adopted reference fields.

a, b, Resulting colour-magnitude diagram (a) based on the field-star sample drawn from the image containing the cluster (b). **c, d,** As a and b,

but for a representative field region taken from a separate image. **e, f,** As c and d, but for a different field region taken from the same, separate image. The black points are stars which are located in the cluster region, whereas the red points are the adopted field stars.

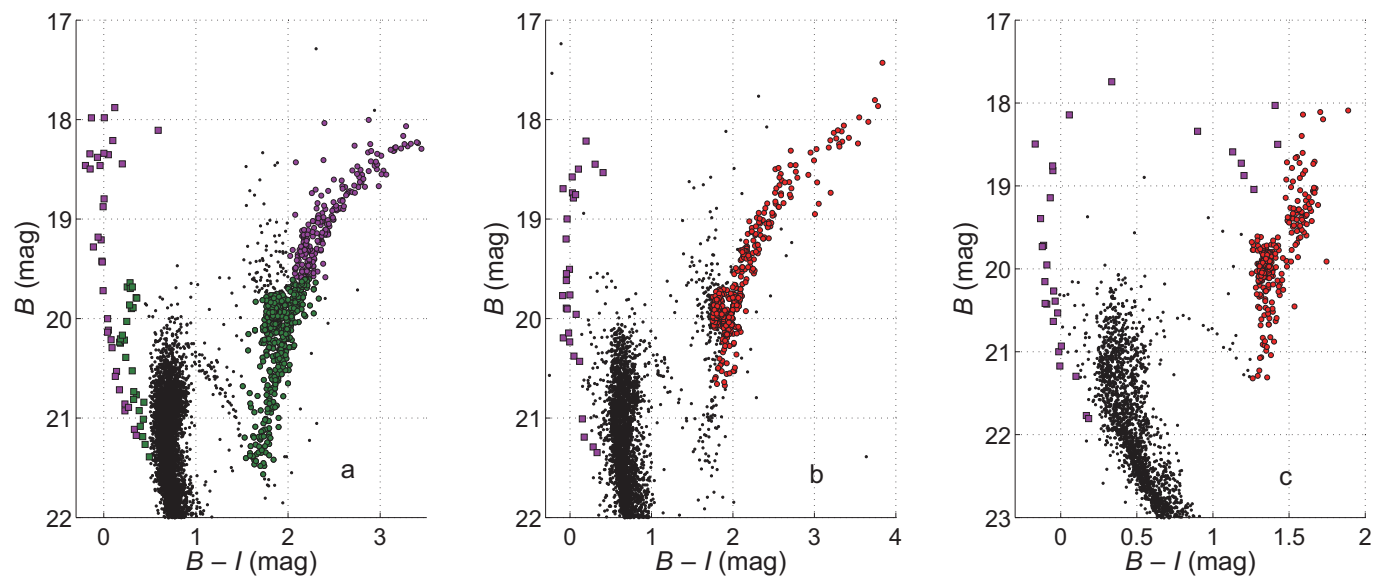


Extended Data Figure 5 | Field-star-decontaminated colour-magnitude diagrams of NGC 1783 for three different grid sizes. **a**, $0.30 \text{ mag} \times 0.15 \text{ mag}$; **b**, $0.40 \text{ mag} \times 0.20 \text{ mag}$; and **c**, $0.50 \text{ mag} \times 0.25 \text{ mag}$.



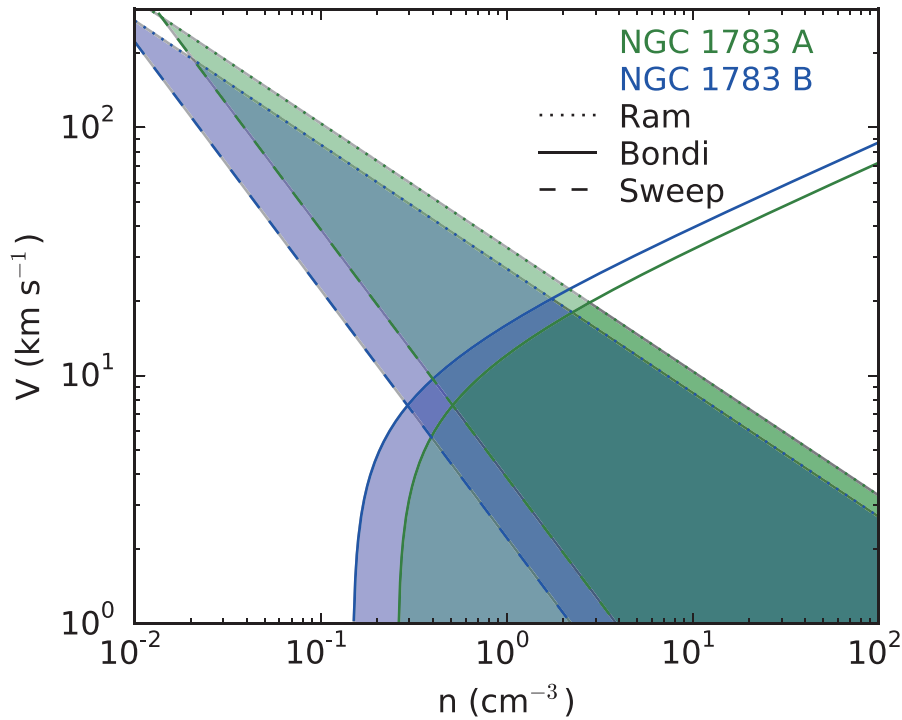
Extended Data Figure 6 | Colour-magnitude analysis. a–d, NGC 1783; e–h, NGC 1696; i–l, NGC 411. First column (a, e, i), raw colour-magnitude diagrams; second column (b, f, j), field-decontaminated colour-magnitude diagrams (also shown in Fig. 1); third column (c, g, k),

colour-magnitude diagrams of the representative field regions; and fourth column (d, h, l), stellar colour-magnitude distributions that were removed from the raw catalogues.



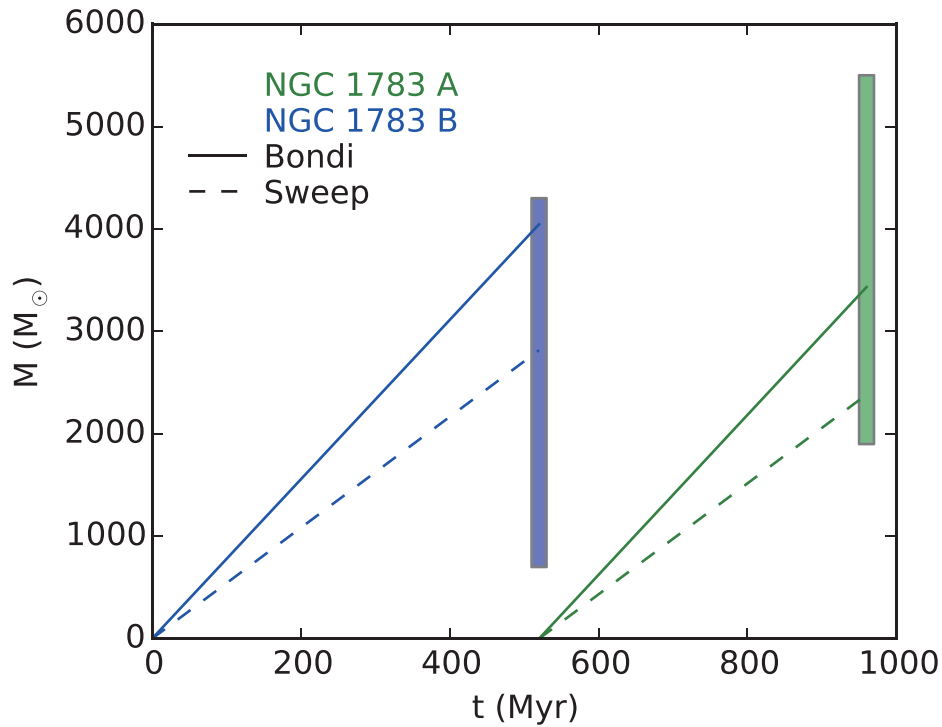
Extended Data Figure 7 | Colour-magnitude diagrams highlighting specific features. **a**, Purple and dark green squares, stars in NGC 1783 sequences A and B, respectively. Dark green circles, corresponding red-giant-branch and red-clump stars, used for comparison with sequence B.

The combination of dark green and purple circles represents the sample used for comparison with sequence A. **b**, Purple squares, NGC 1696 young-sequence stars; red circles, corresponding red-giant-branch and red-clump stars used for comparison. **c**, As **b**, but for NGC 411.



Extended Data Figure 8 | Gas-accretion diagnostic diagram. V is the relative velocity of the cluster with respect to the gas, and n represents the gas density. The shaded regions indicate the parameter space where an NGC 1783-like cluster can accrete the required mass to form the two additional generations of stars, namely one of $250M_{\odot}$ over 520 Myr (blue, corresponding to sequence B in Fig. 1), and a second of $370M_{\odot}$ over 440 Myr (green, corresponding to sequence A in Fig. 1). We have assumed a star-formation efficiency of 10% for all calculations in this figure. The

regions to the right of the 'Gravity' curves correspond to where Bondi accretion can accumulate at least the required mass, and the regions to the right of the 'Sweep' curves correspond to where accretion by collisional sweeping up of ambient interstellar gas by seed intracluster gas can accumulate at least the required mass. The parameter space above the 'Ram' curves are excluded because ram pressure strips clusters of their gas in those regions. See Methods for details.



Extended Data Figure 9 | Gas mass, M , accreted from the interstellar medium as a function of time, t . We have adopted a star-formation efficiency of 10% and calculated representative interstellar gas-accretion frameworks that can explain the stellar masses in the secondary sequences A and B in NGC 1783. For the Bondi regime (solid lines), we used a relative velocity of 4 km s^{-1} and a density of 0.3 cm^{-3} . For the sweeping regime (dashed lines), we used a velocity of 50 km s^{-1} and a density of 0.05 cm^{-3} . The blue and green filled regions indicate the stellar masses

and age offsets of NGC1783 sequences A and B, respectively. The vertical extent of each region provides an estimate of the range in allowed masses. Specifically, for each sequence we plot a region centred on the mass derived using the Kroupa initial mass function, as given in the main text; the range to higher and lower masses is equal to the mean of the differences between the total masses derived using a Salpeter and Kroupa initial mass function of $180 M_{\odot}$ (multiplied by the assumed 10% star formation efficiency).

Measurement noise 100 times lower than the quantum–projection limit using entangled atoms

Onur Hosten¹, Nils J. Engelsen¹, Rajiv Krishnakumar¹ & Mark A. Kasevich¹

Quantum metrology uses quantum entanglement—correlations in the properties of microscopic systems—to improve the statistical precision of physical measurements¹. When measuring a signal, such as the phase shift of a light beam or an atomic state, a prominent limitation to achievable precision arises from the noise associated with the counting of uncorrelated probe particles. This noise, commonly referred to as shot noise or projection noise, gives rise to the standard quantum limit (SQL) to phase resolution. However, it can be mitigated down to the fundamental Heisenberg limit by entangling the probe particles. Despite considerable experimental progress in a variety of physical systems, a question that persists is whether these methods can achieve performance levels that compare favourably with optimized conventional (non-entangled) systems. Here we demonstrate an approach that achieves unprecedented levels of metrological improvement using half a million ⁸⁷Rb atoms in their ‘clock’ states. The ensemble is 20.1 ± 0.3 decibels (100-fold) spin-squeezed via an optical-cavity-based measurement. We directly resolve small microwave-induced rotations 18.5 ± 0.3 decibels (70-fold) beyond the SQL. The single-shot phase resolution of 147 microradians achieved by the apparatus is better than that achieved by the best engineered cold atom sensors despite lower atom numbers^{2,3}. We infer entanglement of more than 680 ± 35 particles in the atomic ensemble. Applications include atomic clocks⁴, inertial sensors⁵, and fundamental physics experiments such as tests of general relativity⁶ or searches for electron electric dipole moment⁷. To this end, we demonstrate an atomic clock measurement with a quantum enhancement of 10.5 ± 0.3 decibels (11-fold), limited by the phase noise of our microwave source.

Quantum noise arises from the impossibility of simultaneously measuring conjugate physical observables. For example, for a harmonic oscillator, capturing the basic physics of light or superconducting circuits, the conjugate observables would be position and momentum. These observables exhibit finite uncertainties even in the lowest possible energy state. For an ensemble of N two-level atoms, a convenient mapping that captures the physics is the pseudo-spin system, where each atom is a spin-half system and the ensemble constitutes a spin- J system ($J = N/2$). The conjugate observables are the Cartesian spin components $J_{x,y,z}$ with the uncertainty relation $\Delta J_x \cdot \Delta J_y \geq |J_z|/2$, and can be pictorially represented on a Bloch sphere (see below). For an unentangled ensemble in a coherent spin state (CSS) with $\langle J_y \rangle = \langle J_z \rangle = 0$ the uncertainties are given by $\Delta J_z = \Delta J_y = \sqrt{N}/2$. This quantity is the projection noise of a CSS (CSS noise).

‘Squeezing’⁸ serves to redistribute the noise to make the observable of interest quieter than the CSS noise level, while still conforming to uncertainty relations. This process is non-classical as it introduces quantum entanglement into the system. The metrological improvement provided by squeezing is quantified by $\chi^2 = \left(\frac{\sqrt{N}/2}{\Delta J_z} \cdot \frac{|J_x|}{N/2} \right)^2$; the first factor represents noise reduction, whereas the second represents coherence loss. This quantity translates directly into reduction

in resources needed to perform a specific measurement (20 dB ($\chi^2 = 100$) is equivalent to a 100-fold increase in atom number or reduction in averaging time).

In optical settings¹⁰ and superconducting microwave circuits¹¹, squeezing in excess of 12 dB has been demonstrated. Implementations in the interferometers of the GEO-600 and LIGO gravitational wave detectors achieved 2.5 dB improvements¹². Other demonstrations include non-invasive biological imaging¹³. Spin-squeezing has been shown in cold atomic ensembles with methods based on both interaction^{14–16} and measurement^{17–19}. States squeezed by 5.6 dB (ref. 20) were used to obtain atomic-clock improvements up to 4.5 dB (ref. 21), and magnetometer enhancements up to 3.4 dB were observed in other experiments^{22,23}. A metrological improvement of 10 dB was attained in a cavity-based experiment²⁴ taking advantage of cycling transitions in ⁸⁷Rb. However, as it utilizes magnetically sensitive states, and atoms are non-uniformly coupled to the cavity, this last approach is not suitable

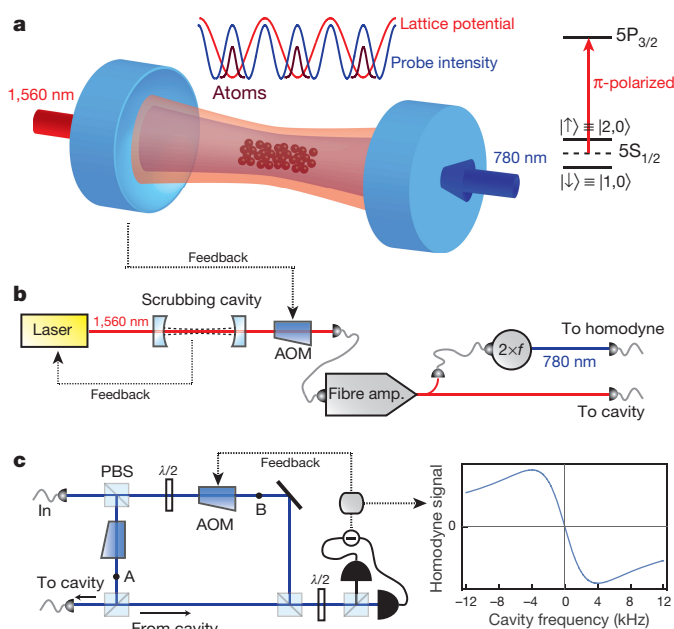


Figure 1 | Overall setup. **a**, Uniform atom–probe coupling: atoms are trapped at the maxima of the probe intensity profile by the 1,560 nm lattice (left). The 780 nm probe light is detuned by equal and opposite amounts from the two clock states (right). **b**, Probe light is generated by frequency-doubling the 1,560 nm light whose frequency is stabilized to match that of the main cavity resonance via feedback control. No further frequency stabilization of the probe is required, eliminating residual 780 nm light inside the cavity. AOM, acousto-optic modulator; PBS, polarizing beam-splitter. **c**, Homodyne detection system and the form of the output signal. Path ‘A’ contains two path length stabilizing side-bands, in addition to the probe frequency, also present on path ‘B’. See Methods for details of **a–c**.

¹Department of Physics, Stanford University, Stanford, California 94305, USA.

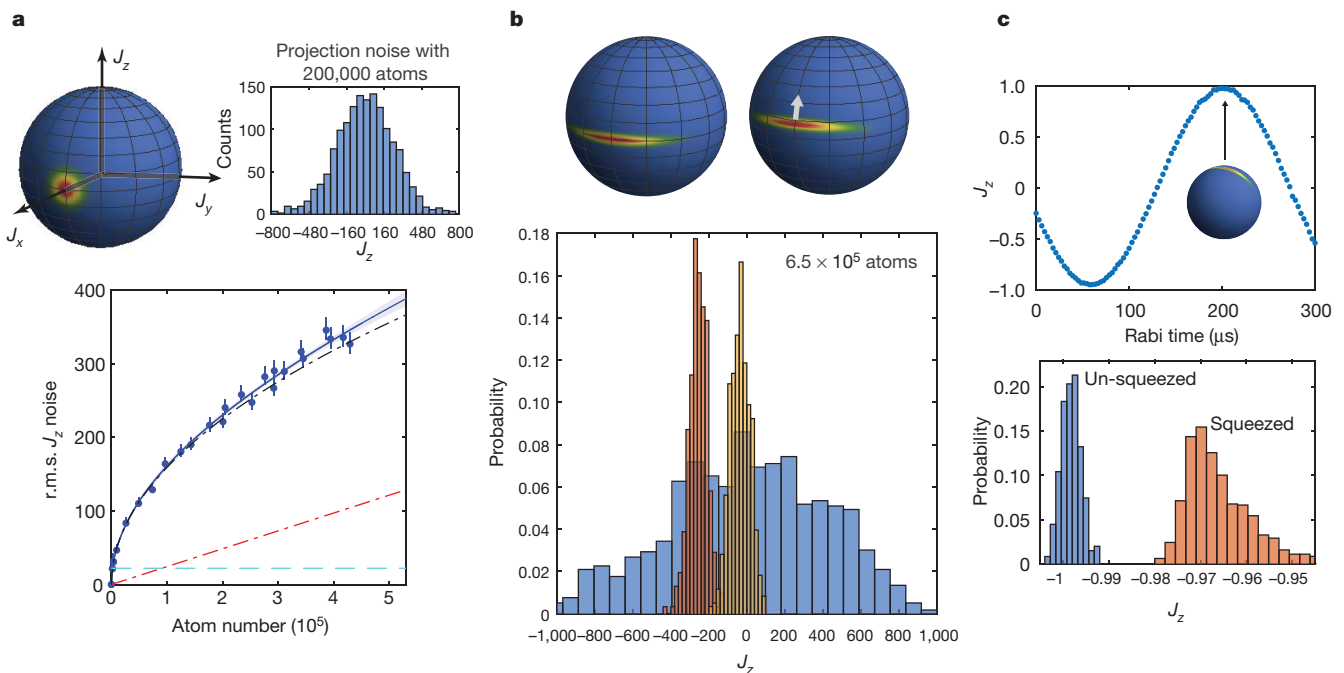


Figure 2 | Squeezing and metrology. **a**, Projection noise of unentangled spins. Top: pictorial representation of a CSS on Bloch sphere (left) and a sample J_z distribution (right). Bottom: characterization of CSS noise. Dot-dashed curve, the expected noise; solid line, fit revealing the underlying microwave rotation noise (dot-dashed straight line); dashed line, resolution of the π -strength measurement (subtracted in quadrature from the data). Error bars, 68% statistical confidence interval. **b**, Top: pictorial representation of two squeezed spin states, one rotated in the

direction of the white arrow by a weak microwave pulse by $660 \mu\text{rad}$. Bottom: the corresponding measured squeezed distributions overlaid on the un-squeezed distribution (measurement strengths, 0.75π – 2.0π). **c**, Top: 96.2%-contrast Rabi oscillations executed by the squeezed state, observed with fluorescence imaging. The Bloch sphere represents the state at the indicated point. Bottom: measured J_z distributions at the lowest point of the Rabi oscillations. The width of the un-squeezed case, shown for reference, is limited by camera read-out noise.

for atomic clocks and other precision sensors requiring the release of trapped atoms from their locations in the cavity.

Owing to systematic errors arising from collisions between atoms, there is typically an upper bound to the number of atoms that can be employed in state-of-the-art cold atom sensors^{2–4}. Squeezing offers a universal path to surpassing this limitation in sensitivity. However, methods demonstrated thus far have fallen significantly short of achieving competitive single-shot phase readout sensitivities.

Here we present a quantum metrology implementation using the magnetically insensitive clock states ($|\uparrow\rangle$ and $|\downarrow\rangle$) of ^{87}Rb (refs 19–21). We prepare the squeezed states through a collective population difference measurement on the atoms. In the spin language, we make a J_z measurement that projects the quantum state into one with a narrower distribution of J_z than that of a CSS. The measurement is enabled by a high-finesse optical cavity: the $|\downarrow\rangle$ ($|\uparrow\rangle$) atoms increase (decrease) the index of refraction seen by the probe light. Therefore, the frequency shift of the cavity resonance is a direct predictor of J_z .

The J_z measurement resolution is determined by the competition between photon shot noise and probe induced Raman scattering (spin-flips). The former limits the precision of the cavity frequency measurements; the latter leads to a random walk in the measured observable. For the ^{87}Rb clock states we expect an optimal enhancement of $\chi_{\text{opt}}^2 \approx \frac{\sqrt{3/2} \varepsilon N C}{1 + N C (\Gamma/\omega_{\text{HF}})^2}$ (see Methods and ref. 25). Here, C is the cooperativity (photon-scattering-rate ratio into cavity mode versus free-space), $\Gamma/\omega_{\text{HF}}$ is the ratio between the excited state linewidth and the hyperfine splitting, and ε is the overall probe detection efficiency. The enhancement saturates with atom number owing to the impact of atomic absorption on measurement sensitivity, in addition to the spin-flips. For our parameters ($\varepsilon = 0.16$), the upper bound on achievable squeezing is about 24 dB.

We operate in a configuration where the probe light is uniformly coupled to all the atoms that are confined in a one-dimensional optical

lattice (Fig. 1a). As we will elaborate, this uniform coupling will enable retrieval of squeezing even if the atoms are released from the optical lattice—a key requirement for many cold atom sensors. In contrast to earlier work^{19–21,24}, we prepare squeezed states without resorting to spin-echo techniques required for non-uniformly coupled systems, and we base squeezing levels on the true CSS noise level of N atoms (instead of an inferred level using lower effective numbers of atoms).

The core apparatus is described in ref. 26. We load the 520- μK -deep optical lattice with a 25 μK ensemble of up to 7×10^5 atoms prepared in the lower clock state (see Methods). Our 10.7-cm near-confocal cavity with a finesse of 1.75×10^5 (linewidth $\kappa_0 = 8.0 \text{ kHz}$) yields $C = 0.78$. The cavity resonance frequency is measured with a homodyne detection system (Fig. 1b, c) that is limited by photon shot noise down to 10 Hz noise frequencies, well below the frequency range of interest (~ 0.2 – 4 kHz). The form of the signal is shown in Fig. 1c. The probe laser frequency is stable down to cavity shifts induced by three spin-flips. For reference, a spin-flip causes a 5.5 Hz shift, a value calculated from the well-known atom–cavity parameters, and verified experimentally at the 10% level with ac-Stark shift measurements for a known intra-cavity probe power.

We first calibrate our CSS noise level. With a $\pi/2$ microwave pulse we bring the state to the equator on the Bloch sphere. To suppress microwave pulse amplitude noise, the implementation is in two steps: $\pi/2_0 - \pi/2_{2\pi/3}$ (subscripts indicate relative phase between pulses). Subsequently, we probe the cavity with a 200 μs probe pulse (see Methods). We obtain a distribution of cavity shifts revealing the J_z distribution (Fig. 2a). Since we measure a balanced population difference, atom number fluctuations ($\sim 2\%$ r.m.s.) between different runs do not enter at a measurable level. Microwave rotation noise becomes noticeable towards 5×10^5 atoms.

Next, we show that the J_z measurements indeed prepare a state with reduced J_z noise. Irrespective of the noise on the first measurement, a second measurement should be correlated (to within the measurement strength) with the first one. We quantify the measurement strength

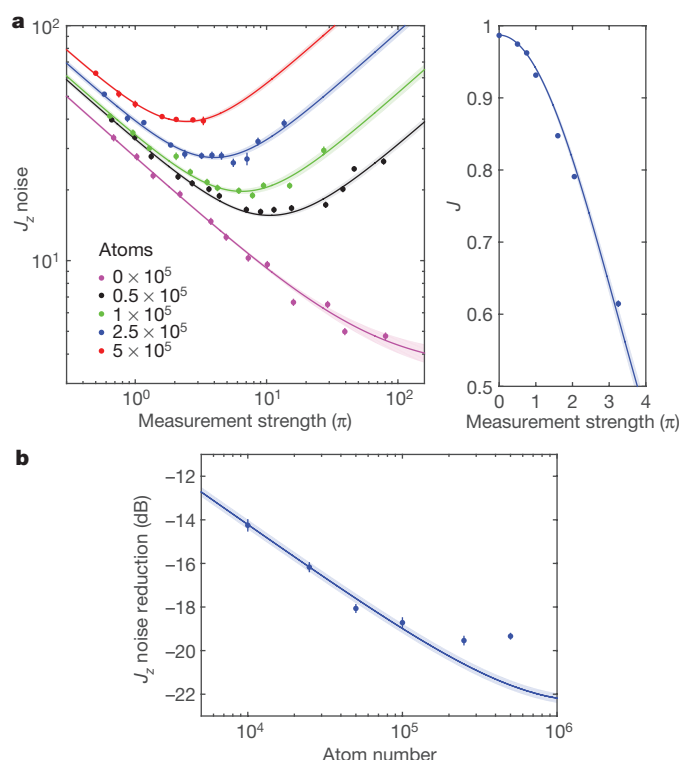


Figure 3 | Measured spin-noise reduction and coherence. **a**, Left: observed spin-noise (r.m.s.) for the difference between two equal-strength back-to-back measurements separated by 1.1 ms (half transverse oscillation period). No-atom data signifies the noise floor. Solid lines, model fits (see Methods). Right: Mean length J of the Bloch vector (coherence), obtained by measuring Ramsey fringe contrasts following the first measurement. Loss in coherence is due to lattice induced ac-Stark shifts which are partially cancelled by probe ac-Stark shifts. Solid line, Gaussian decay fit. **b**, Atom number dependence of maximum observed spin-noise reduction with respect to CSS noise (0 dB is no reduction). Solid line, model fit (see Methods). Error bars and shaded regions in all panels, 68% statistical confidence interval for data and fits respectively.

by the amount of differential phase shift (in radians) accumulated on the clock states due to probe-induced ac-Stark shifts. At high atom numbers, CSS noise exceeds the linear region of the homodyne signal. Outside this region both the measurement efficiency and the intra-cavity probe power decrease; the former degrades achievable squeezing, and the latter results in varying ac-Stark shifts. To remedy this problem, we start with states deterministically pre-squeezed up to 7 dB with 99% coherence (see Methods). Figure 3a shows the noise in the difference between two identical strength back-to-back measurements, as well as coherences, as a function of measurement strength (determined by incident probe power) for different atom numbers. Spin noise reductions at optimal measurement strengths saturate with atom number (Fig. 3b), albeit slightly earlier than expected, suggesting unknown sources of additional noise. States prepared by the first measurement are input states to any subsequent metrology experiment. Owing to the additional uncorrelated noise from the second measurement, these states contain 3 dB more squeezing than directly observed in the measurement difference (see ref. 19). The input state with the largest inferred metrological enhancement is at 5×10^5 atoms with a π measurement strength, which gives 20.1(3) dB (here the digit in parentheses represents the uncertainty in the final digit of the value) enhancement capability including the 0.6 dB loss from the measured 93.2% coherence. In Methods, we show that some of the prepared states contain in excess of 680(35) particle entanglement (following refs 27, 28), and also discuss the level of anti-squeezing in J_y .

To demonstrate a metrology example, we decrease the first measurement strength to 0.75π while increasing the second one to 2.0π .

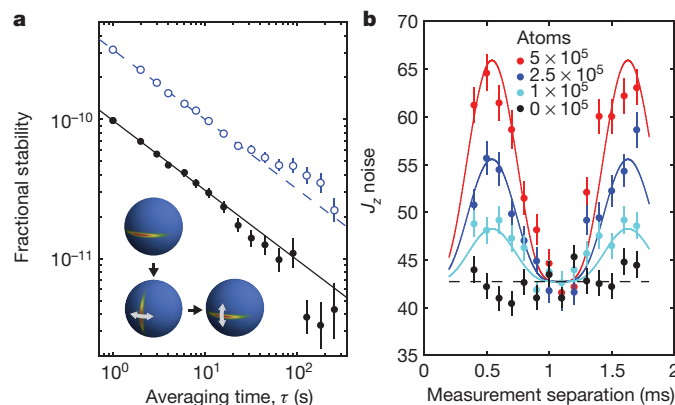


Figure 4 | Clock implementation. **a**, Allan deviation—a measure quantifying clock stability—of a squeezed atomic clock for 228 μ s interrogation time at 1 Hz repetition rate (filled circles). The state is rotated into phase-sensitive orientation during the interrogation time. Dashed line, theoretical CSS noise limit; open circles, measured CSS noise level for the same Ramsey time. Solid line, $9.7 \times 10^{-11} \text{ s}^{1/2}/\sqrt{\tau}$. The Bloch spheres illustrate the clock sequence with squeezed states; white arrows depict conversion of phase jitter into population jitter. **b**, Inhomogeneity analysis: back-to-back measurement noise as a function of measurement separation time. No matching pattern in mean J_z difference between the measurements is observed; only the noise is modulated. Measurement strengths are chosen to make the noise minima equal for different atom numbers. Solid lines are predictions with no free parameters. Error bars in both panels, 68% statistical confidence interval.

In this configuration, coherence after the first measurement is 96.2% (with 2×10^{-3} photons per atom free-space scattering), and the second measurement recovers more information. We obtain a directly measured metrological improvement of 18.5(3) dB with respect to the SQL, with which we resolve small rotations around J_y at 6.5×10^5 atoms (Fig. 2b, c). This corresponds to 147 μ rad (r.m.s.) single-shot phase sensitivity.

To take full advantage of the achieved sensitivity for implementing an atomic clock, we need to reduce the phase noise of our microwave local oscillator. Nevertheless, here we include a preliminary clock demonstration. In particular, we show that we can compare the atomic and microwave phases to better than that allowed by CSS noise. Following J_z -squeezing, we apply a $\pi/2$ pulse (74 μ s) to rotate the noise ellipse into a phase-sensitive state, then wait for phase accumulation, and finally map this phase onto J_z with another rotation (Fig. 4a inset). High-frequency microwave phase noise during the measurement interrogation period results in excess noise proportional to the atom number. To obtain the maximum quantum enhancement, we lower the atom numbers to 1×10^5 and observe up to 10.5(3) dB metrological gain in phase comparison. To put these measurements in context with clock performance²¹, we operate with the largest Ramsey time ($\sim 228 \mu$ s phase accumulation) which does not measurably degrade the comparison. We achieve 9.7×10^{-11} fractional stability at 1 s averaging time (Fig. 4a). For the fixed Ramsey time, the squeezed clock reaches a given precision 11.2(8) times faster than possible without squeezing. High-frequency local oscillator phase noise can be circumvented using interleaved clocks²⁹, in which case the full advantage of squeezing can be achieved using adaptive measurements³⁰.

The squeezed states prepared in this work can be released from the confining lattice for precision sensing applications. To infer the retrievable squeezing in a configuration where the second cavity measurement is replaced with fluorescence imaging in free space, we analyse the noise modulations observed in back-to-back cavity measurements (Fig. 4b) as a function of delay time between the measurements. These modulations are evidently due to residual atom–probe coupling inhomogeneities, and on the basis of our model, we anticipate retrieving up to 14.6 dB squeezing (see Methods).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 June; accepted 22 October 2015.

Published online 11 January 2016.

- Giovannetti, V., Lloyd, S. & Maccone, L. Quantum-enhanced measurements: beating the standard quantum limit. *Science* **306**, 1330–1336 (2004).
- Guéna, J., Abgrall, M., Clairon, A. & Bize, S. Contributing to TAI with a secondary representation of the SI second. *Metrologia* **51**, 108–120 (2014).
- Rocco, E. *et al.* Fluorescence detection at the atom shot noise limit for atom interferometry. *New J. Phys.* **16**, 093046 (2014).
- Bloom, B. J. *et al.* An optical lattice clock with accuracy and stability at the 10^{-18} level. *Nature* **506**, 71–75 (2014).
- Geiger, R. *et al.* Detecting inertial effects with airborne matter-wave interferometry. *Nature Commun.* **2**, 474 (2011).
- Gaaloul, N. *et al.* Quantum tests of the equivalence principle with atom interferometry. *Acta Astronaut.* **67**, 1059–1062 (2010).
- Weiss, D. S., Fang, F. & Chen, J. Measuring the electric dipole moment of Cs and Rb in an optical lattice. *Bull. Am. Phys. Soc.* **APR03**, J1.008 (2003).
- Kitagawa, M. & Ueda, M. Squeezed spin states. *Phys. Rev. A* **47**, 5138–5143 (1993).
- Wineland, D. J., Bollinger, J. J., Itano, W. M. & Heinzen, D. J. Squeezed atomic states and projection noise in spectroscopy. *Phys. Rev. A* **50**, 67–88 (1994).
- Eberle, T. *et al.* Quantum enhancement of the zero-area Sagnac interferometer topology for gravitational wave detection. *Phys. Rev. Lett.* **104**, 251102 (2010).
- Eichler, C., Salathe, Y., Mlynek, J., Schmidt, S. & Wallraff, A. Quantum-limited amplification and entanglement in coupled nonlinear resonators. *Phys. Rev. Lett.* **113**, 110502 (2014).
- The LIGO Scientific Collaboration. A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Phys.* **7**, 962–965 (2011).
- Taylor, M. A. *et al.* Biological measurement beyond the quantum limit. *Nature Photon.* **7**, 229–233 (2013).
- Gross, C., Zibold, T., Nicklas, E., Esteve, J. & Oberthaler, M. K. Nonlinear atom interferometer surpasses classical precision limit. *Nature* **464**, 1165–1169 (2010).
- Riedel, M. F. *et al.* Atom-chip-based generation of entanglement for quantum metrology. *Nature* **464**, 1170–1173 (2010).
- Hamley, C. D., Gerving, C. S., Hoang, T. M., Bookjans, E. M. & Chapman, M. S. Spin-nematic squeezed vacuum in a quantum gas. *Nature Phys.* **8**, 305–308 (2012).
- Appel, J. *et al.* Mesoscopic atomic entanglement for precision measurements beyond the standard quantum limit. *Proc. Natl Acad. Sci. USA* **106**, 10960–10965 (2009).
- Takano, T., Fuyama, M., Namiki, R. & Takahashi, Y. Spin squeezing of a cold atomic ensemble with the nuclear spin of one-half. *Phys. Rev. Lett.* **102**, 033601 (2009).
- Schleier-Smith, M. H., Leroux, I. D. & Vuletic, V. States of an ensemble of two-level atoms with reduced quantum uncertainty. *Phys. Rev. Lett.* **104**, 073604 (2010).
- Leroux, I. D., Schleier-Smith, M. H. & Vuletic, V. Implementation of cavity squeezing of a collective atomic spin. *Phys. Rev. Lett.* **104**, 073602 (2010).
- Leroux, I. D., Schleier-Smith, M. H. & Vuletic, V. Orientation-dependent entanglement lifetime in a squeezed atomic clock. *Phys. Rev. Lett.* **104**, 250801 (2010).
- Sewell, R. J., Koschorreck, M., Napolitano, M., Dubost, B. & Mitchell, M. W. Magnetic sensitivity beyond the projection noise limit by spin squeezing. *Phys. Rev. Lett.* **109**, 253605 (2012).
- Muessel, W., Strobel, H., Linnemann, D., Hume, D. B. & Oberthaler, M. K. Scalable spin squeezing for quantum-enhanced magnetometry with Bose-Einstein condensates. *Phys. Rev. Lett.* **113**, 103004 (2014).
- Bohnet, J. G. *et al.* Reduced spin measurement back-action for a phase sensitivity ten times beyond the standard quantum limit. *Nature Photon.* **8**, 731–736 (2014).
- Chen, Z., Bohnet, J. G., Weiner, J. M., Cox, K. C. & Thompson, J. K. Cavity-aided nondemolition measurements for atom counting and spin squeezing. *Phys. Rev. A* **89**, 043837 (2014).
- Lee, J., Vrijsen, G., Teper, I., Hosten, O. & Kasevich, M. Many-atom-cavity QED system with homogeneous atom-cavity coupling. *Opt. Lett.* **39**, 4005–4008 (2014).
- Sørensen, A. S. & Mølmer, K. Entanglement and extreme spin squeezing. *Phys. Rev. Lett.* **86**, 4431–4434 (2001).
- Lücke, B. *et al.* Detecting multiparticle entanglement of Dicke states. *Phys. Rev. Lett.* **112**, 155304 (2014).
- Biedermann, G. W. *et al.* Zero-dead-time operation of interleaved atomic clocks. *Phys. Rev. Lett.* **111**, 170802 (2013).
- Borregaard, J. & Sørensen, A. S. Near-Heisenberg-limited atomic clocks in the presence of decoherence. *Phys. Rev. Lett.* **111**, 090801 (2013).

Acknowledgements We thank I. Teper, G. Vrijsen and J. Lee for technical contributions to the experiment. This work was supported by DTRA, an NSSEFF fellowship and the ONR.

Author Contributions O.H., N.J.E., R.K. and M.A.K. carried out the experiment, analysed the data, and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.K. (kasevich@stanford.edu).

METHODS

Squeezing measurement details. The incident probe laser power is always increased and decreased adiabatically. The duration of the measurement pulses is 200 μ s—the shortest possible time while avoiding significant ringing in the homodyne signal. Typical incident probe powers are of the order of 10 pW (intra-cavity photon numbers: 90 photons per pW on resonance). Since the homodyne signal is time dependent, we apply a time-dependent weighting function in the analysis of the time traces to extract the cavity frequency shift. This method gives the same signal and photon shot noise values irrespective of pulse shapes and duration as long as the pulse area is conserved and the intra-cavity power adiabatically follows the incident power. Quantitatively, given a time-dependent homodyne signal $s(t)$, the cavity shift is $\Delta\nu = D \int dt s(t) \rho(t) / \int dt \rho^2(t)$, where D is the frequency discriminator (Hz/V) at the peak signal level, and $\rho(t)$ is the temporal shape of the signal normalized such that its peak is at 1. $\rho(t)$ and D are determined experimentally in the absence of atoms by detuning the probe from the exact cavity resonance by a small amount (well within cavity linewidth), and recording the homodyne signal.

The procedure above extracts the correct cavity shift as long as the homodyne signal is in the linear regime. We apply an additional correction factor to properly measure the shifts which are outside this linear regime. This factor is calibrated by setting the cavity–probe detuning to a known value and noting the discrepancy between the known shift and the inferred shift. This is important when establishing CSS noise levels.

At large atom numbers the cavity linewidth broadens owing to atomic scattering, causing the squeezing saturation. The atom-number-dependent linewidth is $\kappa = \kappa_0 + \kappa_S$, with κ_0 the empty cavity linewidth and κ_S the additional scattering contribution. The fractional change is $\kappa_S/\kappa_0 = NC(\Gamma/\omega_{\text{HF}})^2$. We incorporate this signal degrading effect into the cavity shift analysis. The change in the overall signal shape due to broadening is negligible for our parameter range. We experimentally verify that we indeed obtain $\kappa_S/\kappa_0 \approx 0.30$ at 5×10^5 atoms. To do this, we jump the cavity frequency by a known amount (smaller than κ) between consecutive measurements and observe the signal reduction in comparison to the empty cavity case. This measurement takes 10 min; it would have taken 11 h to reach the same precision had we not employed squeezed states.

Homodyne detection system. The homodyne detection setup (Fig. 1c) is seeded with 780 nm light obtained by frequency doubling the 1,560 nm lattice light, which is frequency stabilized to the main cavity following an intermediate stabilization step involving a scrubbing cavity (to narrow down the linewidth of the laser). Thus, the 780 nm light is already stable in the short-term with respect to the cavity (80 mHz/ $\sqrt{\text{Hz}}$ level in the 0.2–4 kHz band—about 8 Hz (r.m.s.) stability). In the long term, thermal drifts in the cavity mirror coatings cause variations in the individual cavity lengths seen by the 780 nm and 1,560 nm light. During experimental cycles of 1 s, the probe frequency can drift by 100 Hz (r.m.s.); we correct the drift at the end of every cycle with an auxiliary empty cavity measurement.

Using a 200 μ W local oscillator (shifted by 80 MHz from input) on path A of the homodyne system, the balanced detectors operate photon shot-noise limited from 10 Hz up to 5 MHz. Two 10 nW spectral components on path B (offset by 78 MHz and 82 MHz from input) travel to the cavity and promptly reflect back from the first mirror to give a heterodyne beat-note signal (2 MHz) at the detectors. This signal is used to stabilize the interferometer path lengths via feedback onto the AOM on path A. Stabilization covers the DC to 15 kHz frequency band, thus removing the influence of optical phase noise for the squeezing measurements. Path B also contains the probe (80 MHz offset from input), interfering with the local oscillator to form the homodyne signal after returning from the cavity.

The overall detection efficiency limiting the achievable squeezing is $\varepsilon = 0.16$. The breakdown is as follows: a factor of 0.50, since we collect light only from one cavity mirror; 0.57 due to loss in cavity mirrors; 0.80 backwards fibre coupling efficiency; 0.85, loss in isolator and other optical elements; 0.85, interferometer mode matching efficiency. Multiplying these factors results in the stated overall efficiency.

Our ability to estimate the centroid of the cavity resonance frequency improves with the square root of the number of photons contained in a 200 μ s measurement pulse, and saturates around 15 Hz owing to laser frequency instability.

Inhomogeneity analysis for free-space release. Although we have uniform atom–cavity coupling, we still have residual inhomogeneities due to thermal motion. Figure 4b shows the noise between back-to-back measurements for varying measurement separations: noise is smallest when the separation is half-integer multiples of the 2.2 ms transverse-oscillation period. We model this behaviour to make an estimate of retrievable squeezing with fluorescence imaging.

The observed effect can be explained with the statistical fluctuations in atomic positions from one experimental cycle to the next. A classical trajectory analysis on the atoms in a harmonic trap (transverse motion) suffices to predict the noise in two consecutive cavity measurements as a function of measurement separation

time. Given their initial phase space points during the first measurement, the location of the atoms during the second measurement, and hence the cavity frequency difference, can be deterministically predicted. However, each time the experiment is repeated, the atoms will start at a different configuration in phase space giving rise to a slightly different cavity frequency difference. We calculate this additional noise assuming a thermal distribution for the atoms. The resulting prediction is

$$\langle \Delta J_z^2 \rangle^{1/2} = \sqrt{N} (1 + \alpha^2) \left(\frac{1}{1 + 2\alpha^2} + \frac{1}{1 + 2\alpha^2 + \alpha^4(1 - \cos^4 \omega t)} - \frac{2}{1 + 2\alpha^2 + \alpha^4 \sin^2 \omega t} \right)^{1/2},$$

where ω is the transverse oscillation angular frequency and $\alpha = 2\sigma_r/w_{780}$ with σ_r the transverse size of the atomic cloud and w_{780} the probe beam waist ($\alpha^2 = 0.076$). The solid lines in Fig. 4b are the addition of this function in quadrature with the baseline without any free parameters.

A similar analysis could be used to predict the additional noise that will arise when the first measurement is done using the cavity and the second one using fluorescence imaging. The second measurement is insensitive to atomic positions; hence the noise comes only from the statistical fluctuations in atomic positions during the first measurement. The estimate for the additional noise is $\langle \Delta J_z^2 \rangle^{1/2} = \sqrt{N} \alpha^2 / \sqrt{1 + 2\alpha^2}$, which is 17 dB below CSS noise for our experiment. In particular, an 18.5 dB squeezed state would read 14.6 dB with fluorescence imaging.

Atom/cavity parameters. For transitions from the clock states with π -polarized light, the single atom cooperativity is $C = 4g^2/\kappa_0\Gamma = 0.78$, with atom–cavity coupling $g = 2\pi \times 96.7$ kHz, empty cavity decay rate $\kappa_0 = 2\pi \times 8.0$ kHz, and atomic decay rate $\Gamma = 2\pi \times 6.06$ MHz. The value for g^2 is an average over the position distribution of the atoms inside the lattice due to the finite temperature. The 25 μ K atoms are distributed over about 1,000 lattice sites. The r.m.s. atomic cloud size inside each 520- μ K-deep lattice site is 17 μ m in the transverse direction and 37 nm in the axial direction. The trap frequencies in the corresponding directions are 460 Hz and 205 kHz. We vary the atom number in the experiment by changing the initial magneto-optical trap (MOT) loading time. Strictly speaking, the residual inhomogeneity in atom–cavity coupling due to thermal motion requires one to define effective atom numbers N_{eff} for the purpose of identifying the CSS noise level. This method was adopted in refs 19, 20, 21 and 24, owing to lack of homogeneity in couplings, resulting in $N_{\text{eff}} \approx 0.66N_0$, with N_0 denoting the real atom number. In this work we achieve $N \equiv N_{\text{eff}} \approx 0.995N_0$ (see next section for details). For practical purposes, we do not differentiate between real and effective atoms.

Coupling inhomogeneity. We would like to measure the collective observable $J_z = \sum_{n=1}^{N_0} j_z^{(n)}$, with N_0 the total number of atoms, and $j_z^{(n)} = \frac{1}{2} \sigma_z^{(n)}$ the z -component of the spin operator for atom n . However, because of the residual inhomogeneity in the atom–cavity coupling we measure a slightly different collective observable J'_z that is a weighted sum over $j_z^{(n)}$.

If all atoms coupled identically to the cavity with the per-spin-flip cavity frequency shift for atom n given by $\delta^{(n)} = \delta_0$, the total shift would be $\Delta_0 = \sum_n \delta_0 j_z^{(n)} = \delta_0 J_z$. However, owing to the small fractional deviations ε_n in the coupling constants for the atoms n , the total shift is given by:

$$\Delta = \sum_n \delta^{(n)} j_z^{(n)} = \sum_n \delta_0 (1 - \varepsilon_n) j_z^{(n)} = (\delta_0 Z) \left(\frac{1}{Z} \sum_n (1 - \varepsilon_n) j_z^{(n)} \right) \equiv \delta_{\text{eff}} J'_z$$

Here Z is a normalization constant, and $\delta_{\text{eff}} = \delta_0 Z$ is the effective cavity shift per spin flip. To decide on a normalization, we utilize two properties of J'_z : its maximum $(J'_z)_{\text{max}} = \frac{N_0}{2} \frac{1}{Z} (1 - \varepsilon)_{\text{e}}$ and its projection noise with uncorrelated atoms $\text{var}(J'_z)_{\text{proj}} = \frac{N_0}{4} \frac{1}{Z^2} \langle (1 - \varepsilon)^2 \rangle_{\text{e}}$. Here $\langle \bullet \rangle_{\text{e}}$ indicates an ensemble average. We choose Z such that the statistical condition $\frac{\text{var}(J'_z)}{(J'_z)_{\text{max}}} = \frac{N_0/4}{N_0/2} = \frac{1}{2}$ satisfied by J_z is also satisfied by J'_z . This leads to $Z = \frac{\langle (1 - \varepsilon)^2 \rangle_{\text{e}}}{(1 - \varepsilon)_{\text{e}}}$, and thus:

$$J'_z = \frac{\langle 1 - \varepsilon \rangle_{\text{e}}}{\langle (1 - \varepsilon)^2 \rangle_{\text{e}}} \sum_n (1 - \varepsilon_n) j_z^{(n)}$$

Consequently, one can think of the non-uniformly coupled system of N_0 atoms as a uniformly coupled system of $N \equiv N_{\text{eff}} = N_0 \frac{\langle 1 - \varepsilon \rangle_{\text{e}}^2}{\langle (1 - \varepsilon)^2 \rangle_{\text{e}}} = 0.995N_0$ effective atoms in conjunction with an effective cavity shift $\delta_{\text{eff}} = \delta_0 \frac{\langle (1 - \varepsilon)^2 \rangle_{\text{e}}}{(1 - \varepsilon)_{\text{e}}} = 0.93\delta_0$ per spin-flip. Here δ_0 is the cavity shift for an atom on the cavity axis averaged over the distribution along the tightly trapped longitudinal direction. This gives $\delta_{\text{eff}} = 0.83\delta_{\text{max}} = 5.5$ Hz, where δ_{max} is the cavity shift for an atom localized at a peak of the probe mode profile.

Fluorescence imaging. Using the cavity, we detect J_z only in a very narrow range. For observing J_z in its full range of $\pm N/2$ we use fluorescence imaging. We release

the atoms from the lattice, and for state selectivity, we apply a laser beam resonant with the $F=2$ to $F'=3$ transition to momentarily push the $F=2$ (that is, $|\uparrow\rangle$) atoms. This spatially separates the $|\uparrow\rangle$ and $|\downarrow\rangle$ states, permitting the simultaneous detection of the number of atoms in both states. We image the fluorescence from the two clouds for 2 ms on a CCD camera using the cooling and re-pumping lights of the MOT. This is how the observed Rabi and Ramsey fringes are mapped out and hence how the coherence is measured.

The pre-squeezing procedure. The pre-squeezing procedure is a supplement to the J_z measurement protocol, increasing robustness and efficiency. However we note that it does not alter the nature of J_z measurements—we obtain the same final squeezing results in the absence of pre-squeezing if we post-select the runs with the first measurement outcome lying within the linear regime of the homodyne signal.

When approaching the saturated regime of squeezing, discussed in the main text, the r.m.s. cavity shifts due to CSS noise approach the linewidth of the cavity. This quantum noise prevents us from preparing initial J_z distributions that fall purely within the linear region of the homodyne signal using microwave rotations alone. We therefore resort to atom–cavity nonlinearities to deterministically pre-squeeze the state by a sufficient amount such that the distribution fits within the linear regime. The nonlinearity employed is a J_z^2 interaction causing one-axis twisting, similar to the one used in ref. 20. For an initial state near the J_x axis, the procedure dynamically compresses the distribution on the Bloch sphere in the z direction while expanding it in the y direction. The technical noise on the initial J_z preparation is also suppressed. The pre-squeezing occurs after the composite $\pi/2$ pulse brings the state to the equator. We send 100 nW of light at $6.25\kappa_0$ detuning from the bare cavity resonance (generating J_z^2 interaction) and simultaneously turn on a $400\mu\text{s}$ microwave $\pi/12$ pulse (generating rotations around the J_x axis). With the combined action, we observe up to 7 dB unconditional J_z squeezing while retaining 99% coherence.

Amount of anti-squeezing. The cavity measurements, while projecting the atomic ensemble onto a state with reduced J_z noise, also act back onto the conjugate observable J_y and increase its noise. In the ideal case a measurement would preserve the area of the uncertainty ellipse, that is, the reduction and the increase in the noises of J_z and J_y respectively would be through the same factor. However owing to photon losses, inefficiencies in extracting the information in the read-out, and the additional spin-flip noise in J_z , the balance is expected to be broken. Experimentally, the variance of J_y scales linearly with measurement strength, reaching 39 dB above CSS noise at π measurement strength accompanying the quoted 20.1(3) dB squeezing in J_z . This corresponds to a factor of 8.8 increase in the uncertainty ellipse area. The observed level of anti-squeezing is within 2 dB of the expectations.

Squeezing limits calculation. The maximum attainable spin noise reduction can be found by considering the individual effects of the measurement noise and the spin-flip noise. We will specify these quantities as functions of the (experimentally accessible) differential phase shift $\phi_{AC} = \frac{2g^2}{\Delta} \int dt n_c(t)$ accumulated on the clock states. Here, $n_c(t)$ is the intra-cavity photon number. We will also use the cooperativity, $C = \frac{4g^2}{\kappa_0\Gamma}$.

The number of scattered photons is $m_s = \phi_{AC}(\Gamma/\omega_{HF})$. The hyperfine splitting enters because the atom–cavity detuning is set to $\Delta = \omega_{HF}/2$. Using the branching ratios for ^{87}Rb , it can be shown that only 1/6 of the scattering events will give rise to a spin-flip, that is, a change of the hyperfine state. These spin-flips will give rise to a random walk on J_z with a variance of $\delta_{\text{flip}}^2 = \frac{N}{6} \frac{\Gamma}{\omega_{HF}} \phi_{AC}$. This is the spin-flip noise; it grows with atom number and probe power.

To examine the measurement noise, we analyse the information imprinted on the light transmitted from the cavity. The total decay rate of the cavity is $\kappa = 2\kappa_M + \kappa_L + \kappa_S$, where κ_M is due to mirror out-coupling, and κ_L is due to optical losses in the mirrors. The term due to atomic scattering can be expressed as $\kappa_S/\kappa_0 = NC(\Gamma/\omega_{HF})^2$, where $\kappa_0 = 2\kappa_M + \kappa_L$. Around zero cavity–probe detuning, the number of photons transmitted through the cavity is $n_T = \frac{\varepsilon_c(\omega_{HF}/\Gamma)}{2C} \phi_{AC}$, where $\varepsilon_c = 2\kappa_M/\kappa_0$ is the cavity efficiency incorporating the optical losses. As the atoms shift the cavity frequency by $\frac{2g^2}{\Delta} J_z$, the phase shift on the light upon

transmission is $\varphi = \frac{2C(\Gamma/\omega_{HF})}{1 + NC(\Gamma/\omega_{HF})^2} J_z$. Given the quantum phase noise $\frac{1}{2\sqrt{n_T}}$ for a coherent state, the noise equivalent J_z resolution is $\frac{1 + NC(\Gamma/\omega_{HF})^2}{\sqrt{8\varepsilon_c C}(\Gamma/\omega_{HF})} \frac{1}{\sqrt{\phi_{AC}}}$. For a

symmetric cavity, equal amounts of information leak out from each mirror. Thus, including the information gained from the reflection would improve the resolution by $\sqrt{2}$. Lastly, we also include the effect of photon losses on the way to the detectors, and bundle all efficiency factors into the quantity ε which is further discussed in the ‘Atom/cavity parameters’ section above. The final expression for the measurement noise is $\delta_{\text{meas}}^2 = \frac{(1 + NC(\Gamma/\omega_{HF})^2)^2}{16\varepsilon C(\Gamma/\omega_{HF})} \frac{1}{\phi_{AC}}$; it decreases with increasing probe power.

As a first approximation, the total noise in the estimation of J_z can be found by adding the contribution due to the two sources: $\delta^2 = \delta_{\text{meas}}^2 + \delta_{\text{flip}}^2$. An optimal ϕ_{AC} (that is, measurement strength) minimizes this expression, at which point $\delta^2 = \frac{\sqrt{N}(1 + NC(\Gamma/\omega_{HF})^2)}{\sqrt{24\varepsilon C}}$. Assuming negligible coherence loss, as is the case experimentally, we arrive at an optimal metrological enhancement $\chi_{\text{opt}}^2 = \frac{\sqrt{\varepsilon NC} 3/2}{1 + NC(\Gamma/\omega_{HF})^2}$.

At first, the achievable enhancement increases with atom number, attaining a maximum of $(\omega_{HF}/\Gamma)\sqrt{3\varepsilon/8}$ at $N_{\text{opt}} = \omega_{HF}^2/\Gamma^2 C$. This saturation effect can be traced back to cavity linewidth broadening from atomic absorption. For $\varepsilon = 1$, the maximum achievable enhancement is around 28 dB.

Exact numerical agreement should not be expected between the naive model presented here and the experiment, since the latter is more complicated. We use the functional forms derived here to fit to the data in Fig. 3. In particular, for Fig. 3a we use $\delta^2 = \frac{\alpha}{\phi_{AC}} + \beta\phi_{AC}$, and for Fig. 3b we use $\chi_{\text{opt}}^2 = \gamma \frac{\sqrt{NC}}{1 + NC(\Gamma/\omega_{HF})^2}$. Here, α , β and γ are the fit parameters.

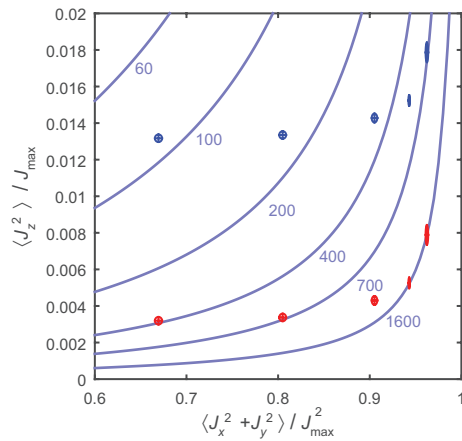
Entanglement depth. Measurements of the collective spin operators $J_i = \sum_{n=1}^N J_i^{(n)}$ for an ensemble of N two-level atoms can be used to quantify the amount of entanglement in the ensemble^{27,28} without further reference to the specific nature of the states. We will follow the analysis of ref. 28, where the J_z variance and the mean-square Bloch vector length places the measured states on a plane with boundaries corresponding to different entanglement depths.

In our experiment, because of the residual inhomogeneities in atom–cavity coupling, we measure the collective observable $J'_z = \frac{1}{Z} \sum_{n=1}^N (1 - \varepsilon_n) J_z^{(n)}$, where Z is a normalization constant and ε_n is the small fractional deviation in coupling for atom n (see ‘Coupling inhomogeneity’ section). This implies that we cannot directly utilize the measured spin noise values for the purposes of calculating entanglement depths. However, even without a direct measurement of J_z itself, its maximum possible variance can be inferred, as was argued in the Methods section ‘Inhomogeneity analysis for free-space release’. There, we found that there would be an additive noise of 17 dB below CSS noise if we tried to read out the prepared states via fluorescence imaging, which we consider to be a true J_z measurement. Therefore, we infer J_z on the basis of this analysis. Unlike the cavity-based measurements, the Bloch vector length measurements, which are done via fluorescence imaging, can directly be used for calculating entanglement depths.

In Extended Data Fig. 1, we plot the inferred J_z variances for the 5×10^5 atom data set using the experimentally established noise after the first measurement ($\frac{1}{\sqrt{2}}$ of the values in Fig. 3a). The point with the largest metrological gain (π measurement strength) gives an entanglement depth of 330(15) atoms, while the largest entanglement depth is 680(35) (0.5π measurement strength) atoms. This exemplifies that entanglement depth is in itself not a direct predictor for metrological improvement.

The additional noise in our model in inferring J_z originates from shot-to-shot randomization of atomic positions. In the absence of this randomization, we expect the discrepancy between the variances of J'_z and J_z to be less. Thus, the quoted entanglement depths should be taken as lower bounds. For reference, had we not taken into account the coupling inhomogeneity we would have found the largest entanglement depth to be 1,605(30) atoms.

Sample size. No statistical methods were used to predetermine sample size.



Extended Data Figure 1 | Inferred entanglement depths, quantifying multi-particle entanglement. The inferred spin noise variance (y axis) and the mean-square Bloch vector lengths (x axis) are plotted for the 5×10^5 atom data set. Note that the probe power decreases from left to right. The x-axis values are conservatively chosen to be the most probable value of the measured Bloch vector length distributions (Fig. 2c). A state below an M -particle boundary (purple lines labelled with particle numbers) is guaranteed to contain at least groups of M particles whose quantum states are non-separable. The blue data set establishes a lower bound on entanglement depth taking into account the residual inhomogeneity in atom-cavity coupling. The red data set, for reference, shows what we would have obtained had we ignored the small inhomogeneity. The ellipses correspond to the 68% statistical confidence intervals on the quoted values. $J_{\max} = N/2$. The third data point in each set shows the largest metrological improvement.

Fully integrated wearable sensor arrays for multiplexed *in situ* perspiration analysis

Wei Gao^{1,2,3*}, Sam Emaminejad^{1,2,3,4*}, Hnin Yin Yin Nyein^{1,2,3}, Samyuktha Challa⁴, Kevin Chen^{1,2,3}, Austin Peck⁵, Hossain M. Fahad^{1,2,3}, Hiroki Ota^{1,2,3}, Hiroshi Shiraki^{1,2,3}, Daisuke Kiriya^{1,2,3}, Der-Hsien Lien^{1,2,3}, George A. Brooks⁵, Ronald W. Davis⁴ & Ali Javey^{1,2,3}

Wearable sensor technologies are essential to the realization of personalized medicine through continuously monitoring an individual's state of health^{1–12}. Sampling human sweat, which is rich in physiological information¹³, could enable non-invasive monitoring. Previously reported sweat-based and other non-invasive biosensors either can only monitor a single analyte at a time or lack on-site signal processing circuitry and sensor calibration mechanisms for accurate analysis of the physiological state^{14–18}. Given the complexity of sweat secretion, simultaneous and multiplexed screening of target biomarkers is critical and requires full system integration to ensure the accuracy of measurements. Here we present a mechanically flexible and fully integrated (that is, no external analysis is needed) sensor array for multiplexed *in situ* perspiration analysis, which simultaneously and selectively measures sweat metabolites (such as glucose and lactate) and electrolytes (such as sodium and potassium ions), as well as the skin temperature (to calibrate the response of the sensors). Our work bridges the technological gap between signal transduction, conditioning (amplification and filtering), processing and wireless transmission in wearable biosensors by merging plastic-based sensors that interface with the skin with silicon integrated circuits consolidated on a flexible circuit board for complex signal processing. This application could not have been realized using either of these technologies alone owing to their respective inherent limitations. The wearable system is used to measure the detailed sweat profile of human subjects engaged in prolonged indoor and outdoor physical activities, and to make a real-time assessment of the physiological state of the subjects. This platform enables a wide range of personalized diagnostic and physiological monitoring applications.

Wearable electronics are devices that can be worn or mated with human skin to continuously and closely monitor an individual's activities, without interrupting or limiting the user's motions^{1–9}. Thus wearable biosensors could enable real-time continuous monitoring of an individual's physiological biomarkers^{10–12}. At present, commercially available wearable sensors are only capable of tracking an individual's physical activities and vital signs (such as heart rate), and fail to provide insight into the user's health state at molecular levels. Measurements of human sweat could enable such insight, because it contains physiologically and metabolically rich information that can be retrieved non-invasively¹³. Sweat analysis is currently used for applications such as disease diagnosis, drug abuse detection, and athletic performance optimization¹³. For these applications, the sample collection and analysis are performed separately, failing to provide a real-time profile of sweat content secretion, while requiring extensive laboratory analysis using bulky instrumentation¹⁹. Recently, wearable sweat sensors have been developed, with which a variety of biosensors

have been used to measure analytes of interest (Supplementary Table 1)^{14–18}.

Given the multivariate mechanisms that are involved in sweat secretion, an attractive strategy would be to devise a fully integrated multiplexed sensing system to extract the complex information available from sweat. Here we present a wearable flexible integrated sensing array (FISA) for simultaneous and selective screening of a panel of biomarkers in sweat (Fig. 1a). Our solution bridges the existing technological gap between signal transduction (electrical signal generation by sensors), conditioning (here, amplification and filtering), processing (here, calibration and compensation) and wireless transmission in wearable biosensors by merging commercially available integrated-circuit technologies, consolidated on a flexible printed circuit board (FPCB), with flexible and conforming sensor technologies fabricated on plastic substrates. This approach decouples the stringent mechanical requirements at the sensor level and electrical requirements at the signal conditioning, processing and transmission levels, and at the same time exploits the strengths of the underlying technologies. The independent and selective operation of individual sensors is preserved during multiplexed measurements by employing highly specific surface chemistries and by electrically decoupling the operating points of each sensor's interface. This platform is a powerful tool with which to advance large-scale and real-time physiological and clinical studies by facilitating the identification of informative biomarkers in sweat.

As illustrated in Fig. 1a, the FISA allows simultaneous and selective measurement of a panel of metabolites and electrolytes in human perspiration as well as skin temperature during prolonged indoor and outdoor physical activities. By fabricating the sensors on a mechanically flexible polyethylene terephthalate (PET) substrate, a stable sensor-skin contact is formed, while the FPCB technology is exploited to incorporate the critical signal conditioning, processing, and wireless transmission functionalities using readily available integrated-circuit components (Fig. 1b). The panel of target analytes and skin temperature was selected to facilitate an understanding of an individual's physiological state (see Supplementary Information for selection of the target analytes). For example, excessive loss of sodium and potassium in sweat could result in hyponatremia, hypokalemia, muscle cramps or dehydration²⁰; sweat glucose is reported to be metabolically related to blood glucose²¹; sweat lactate can potentially serve as a sensitive marker of pressure ischaemia²²; and skin temperature is clinically informative of a variety of diseases and skin injuries such as pressure ulcers^{23,24}. Additionally, skin temperature measurements are needed to compensate for and eliminate the influence of temperature variation in the readings of the chemical sensors through a built-in signal processor.

Figure 1c illustrates the schematic of the multiplexed sensor array (each electrode is 3 mm in diameter) for sweat analysis; fabrication processes are detailed in Methods and Extended Data Fig. 1. Here,

¹Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA. ²Berkeley Sensor and Actuator Center, University of California, Berkeley, California 94720, USA. ³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁴Stanford Genome Technology Center, Stanford School of Medicine, Palo Alto, California 94304, USA. ⁵Integrative Biology, University of California, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

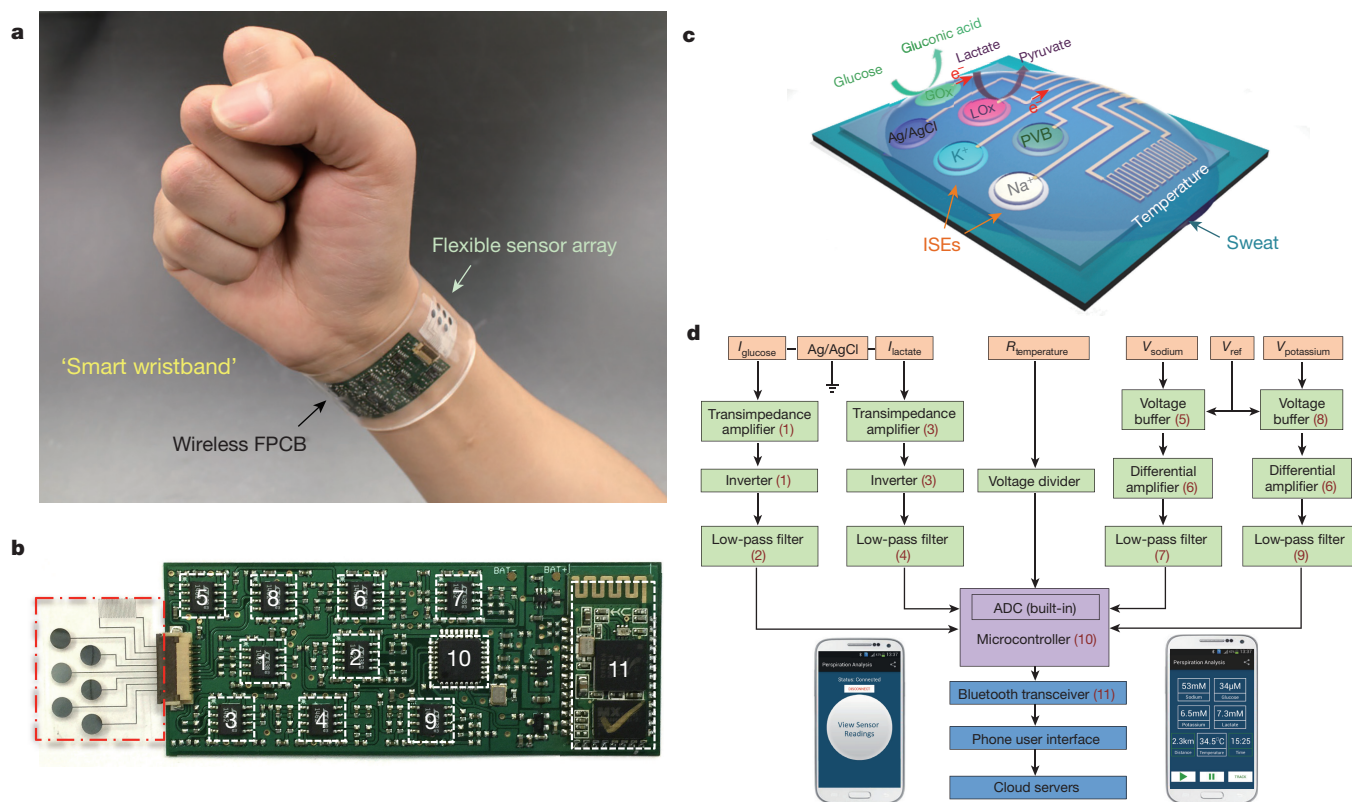


Figure 1 | Images and schematic illustrations of the FISA for multiplexed perspiration analysis. **a**, Photograph of a wearable FISA on a subject's wrist, integrating the multiplexed sweat sensor array and the wireless FPCB. (All photographs in this paper were taken by the authors.) **b**, Photograph of a flattened FISA. The red dashed box indicates the location of the sensor array and the white dashed boxes indicate the locations of the integrated circuit components. **c**, Schematic of the sensor array (including glucose, lactate, sodium, potassium and temperature sensors) for multiplexed perspiration analysis. GOx and LOx, glucose

amperometric glucose and lactate sensors (with current output) are based on glucose oxidase and lactate oxidase immobilized within a permeable film of the linear polysaccharide chitosan. A Ag/AgCl electrode serves as a shared reference electrode and counter electrode for both sensors. The use of Prussian blue dye as a mediator minimizes the reduction potentials to approximately 0 V (versus Ag/AgCl) (Extended Data Fig. 2a), and thus eliminates the need for an external power source to activate the sensors. These enzymatic sensors autonomously generate current signals proportional to the abundance of the corresponding metabolites between the working electrode and the Ag/AgCl electrode. The measurement of Na^+ and K^+ levels is facilitated through the use of ion-selective electrodes (ISEs), coupled with a polyvinyl butyral (PVB)-coated reference electrode to maintain a stable potential in solutions with different ionic strengths (Extended Data Fig. 2b–d). By using poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS) as an ion-to-electron transducer in the ISEs and carbon nanotubes in the PVB reference membrane²⁵, robust potentiometric sensors (with voltage output) can be obtained for long-term continuous measurements with negligible voltage drift. A resistance-based temperature sensor is realized by fabricating Cr/Au metal microwires. Parylene is used as an insulating layer to ensure reliable sensor reading by preventing electrical contact of the metal lines with skin and sweat.

Figure 1d illustrates the system-level overview of the signal transduction, conditioning, processing, and wireless transmission paths to facilitate multiplexed on-body measurements. The signal-conditioning path for each sensor is implemented with analogue circuits and in relation to the corresponding transduced signal. The circuits are configured to ensure that the final analogue output of each

oxidase and lactate oxidase. **d**, System-level block diagram of the FISA showing the signal transduction (orange) (with potential V , current I and resistance R outputs), conditioning (green), processing (purple) and wireless transmission (blue) paths from sensors to the custom-developed mobile application (numbers in parentheses indicate the corresponding labelled components in **b**). ADC, analogue-to-digital converter. The inset images show the home page (left) and the real-time data display page (right) of the mobile application.

path is finely resolved while staying within the input voltage range of the analogue-to-digital converter. Furthermore, the microcontroller's computational and serial communication capabilities are used to calibrate, compensate, and relay the conditioned signals to an on-board wireless transceiver. The transceiver facilitates wireless data transmission to a Bluetooth-enabled mobile handset with a custom-developed application (Extended Data Fig. 3), containing a user-friendly interface for sharing (through email, SMS, and so on) or uploading the data to cloud servers. The circuit design, calibration, and power delivery diagram of the FISA are described in Methods and Extended Data Figs 4 and 5.

The performance of each sensor was monitored separately with different analyte solutions. Figure 2a and b shows the representative current responses of the glucose and lactate sensors, measured chronoamperometrically in 0–200- μM glucose solutions and 0–30-mM lactate solutions, respectively. A linear relationship between current and analyte concentrations with sensitivities of $2.35 \text{ nA } \mu\text{M}^{-1}$ for glucose sensors and 220 nA mM^{-1} for lactate sensors was observed. Figure 2c and d illustrates the open circuit potentials of Na^+ and K^+ sensors in the electrolyte solutions with physiologically relevant concentrations of 10–160-mM Na^+ and 1–32-mM K^+ respectively. Both ion-selective sensors show a near-Nernstian (according to the Nernstian equation, the theoretical sensitivity of the ISE-based sensors should be 59) behaviour with sensitivities of 64.2 mV and 61.3 mV per decade of concentration for Na^+ and K^+ sensors, respectively. Results of repeatability and long-term stability studies indicate that the sensitivities of the biosensors are consistent over a period of at least four weeks (Extended Data Fig. 6). Figure 2e displays the linear response of the resistive temperature

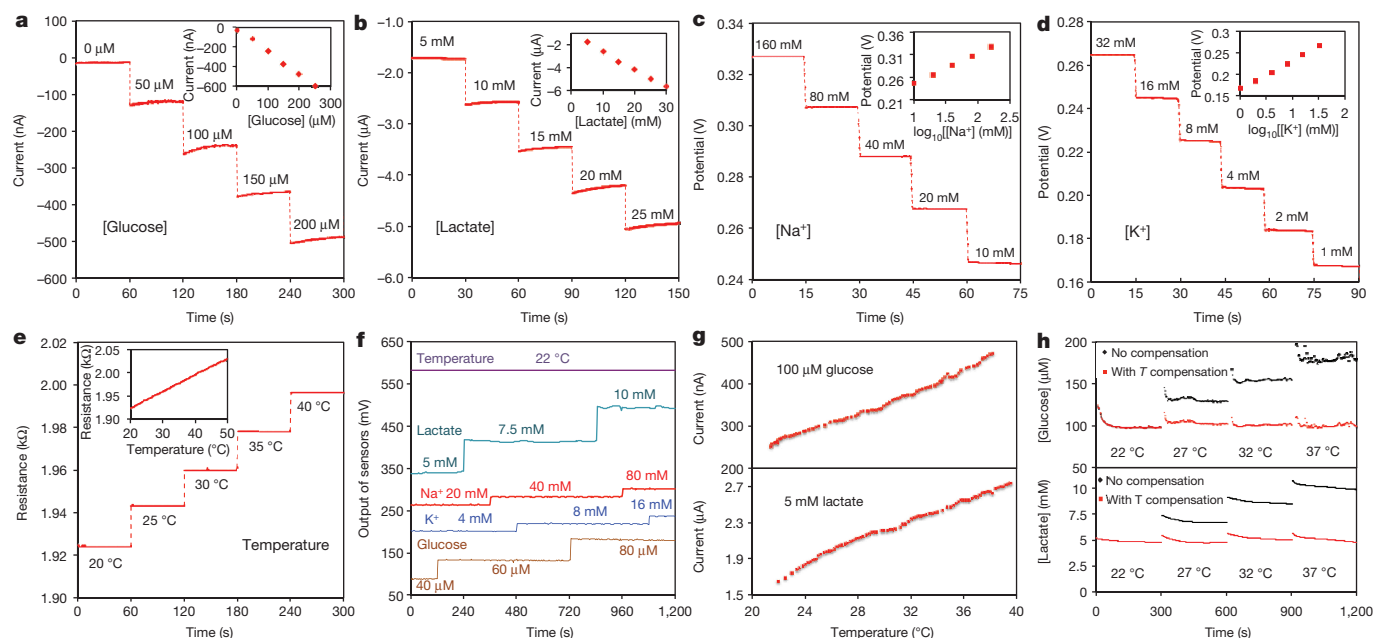


Figure 2 | Experimental characterizations of the wearable sensors. **a, b,** The chronoamperometric responses of the glucose (**a**) and lactate (**b**) sensors to the respective analyte solutions in phosphate-buffered saline (PBS). **c, d,** The open circuit potential responses of the sodium (**c**) and potassium (**d**) sensors in NaCl and KCl solutions. **e,** The resistance response of the temperature sensor to temperature changes (20–40 °C) in

PBS. Insets in **a–e** show the corresponding calibration plots of the sensors. Data recording was paused for 30 s for each solution change in **a–e**. **f,** System-level interference studies of the sensor array. **g,** The influence of temperature on the responses of the glucose and lactate sensors. **h,** System-level real-time temperature T compensation for the glucose and lactate sensors in 100- μ M glucose and 5-mM lactate solutions, respectively.

sensor in the physiological skin temperature range of 20–40 °C with a sensitivity of approximately 0.18% per degree Celsius (normalized to the resistance at 20 °C).

The selectivity of sweat sensors is crucial, because various electrolytes and metabolites in sweat can influence the accuracy of the sensor readings. Extended Data Fig. 7a–d shows that the presence of non-target electrolytes and metabolites causes negligible interference to the response of each sensor. When all five sensors are integrated in the FISA, simultaneous system-level measurements maintain excellent selectivity upon varying concentrations of each analyte (Fig. 2f and Extended Data Fig. 7e–h). Although temperature has a minimal effect on the potentiometric sensors, it greatly influences the performance of the enzymatic sensors. Figure 2g shows that the responses of glucose and lactate sensors increase rapidly upon elevation of the solution temperature from 22 °C to 40 °C, reflecting the effect of increased enzyme activities²⁶. System integration allows for the implementation of real-time compensation to calibrate the sensor readings on the basis of temperature variations. Figure 2h illustrates that with the increase of temperature, the uncompensated sensor readouts can lead to substantial overestimation of the actual concentration of the given glucose and lactate solutions; however, the temperature compensation allows for accurate and consistent readings.

It is essential for wearable devices to be able to withstand the stress of daily human wear and physical exercise. A study on mechanical deformation conducted by monitoring the performance of both the sensor array and the FPCB before, during and after bending (radii of curvature are 1.5 cm and 3 cm, respectively) (Extended Data Fig. 8) reveals minimal output changes in the FISA's responses.

The FISAs can be comfortably worn on various body parts, including the forehead, wrists and arms. Figure 3a shows a human subject wearing two FISAs, packaged as a 'smart wristband' and a 'smart headband', allowing for real-time perspiration monitoring on the wrist and forehead simultaneously during stationary leg cycling. To ensure the fidelity of sensor readings, the data collection of each channel took place when a sufficient sweat sample was present, as shown by stabilization of sensor readings within the physiologically relevant

range (see Methods). The accuracy of on-body measurements was verified through the comparison of on-body sensor readings from the forehead with *ex situ* (off-body) measurements from collected sweat samples (Fig. 3b).

Real-time physiological monitoring was performed on a subject during constant-load exercise on a cycle ergometer. The protocol involved a 3-min ramp-up, 20-min cycling at 150 W, and a 3-min cool-down. During the exercise, the heart rate, oxygen consumption (V_{O_2}), and pulmonary minute ventilation were measured using external monitoring instruments, and were found to increase proportionally with increasing power output as shown in Fig. 3c. Figure 3d illustrates the corresponding real-time measurements on the subject's forehead using a FISA. The skin temperature remains constant at 34 °C up to perspiration initiation at about 320 s. The dip in temperature at this point indicates the beginning of perspiration and evaporative cooling²⁷. With continued perspiration, the skin temperature rises at about 400 s because of muscle heat conductance to skin and then remains stable, while the concentration of both lactate and glucose in sweat decrease gradually. The decreases in concentration of lactate and glucose in sweat are expected, owing to the dilution effect caused by an increase in sweat rate, which is visually observed as exercise continues¹³. However, lactate concentration becomes relatively stable after 1,100 s, indicating the stabilization of physiological responses to continuous, sub-maximal constant-exercise power output²². Sweat $[Na^+]$ increases and $[K^+]$ decreases in the beginning of perspiration, in line with the previous *ex situ* studies from the collected sweat samples^{28,29}. Both $[Na^+]$ and $[K^+]$ stabilize as the cycling continues. By wearing a FISA on different parts of the body, the site-specific variations in electrolyte and metabolite levels²⁹ can also be monitored and studied simultaneously. Sweat analyte levels on the wrist follow similar trends but with concentrations different from those obtained at the forehead (Extended Data Fig. 9). In this case, because the subject had a lower sweat rate at the wrist²⁹, the sensors were activated at a later time.

The physiological response of the subjects to a sudden change in exercise intensity was also investigated, in a graded-load exercise which

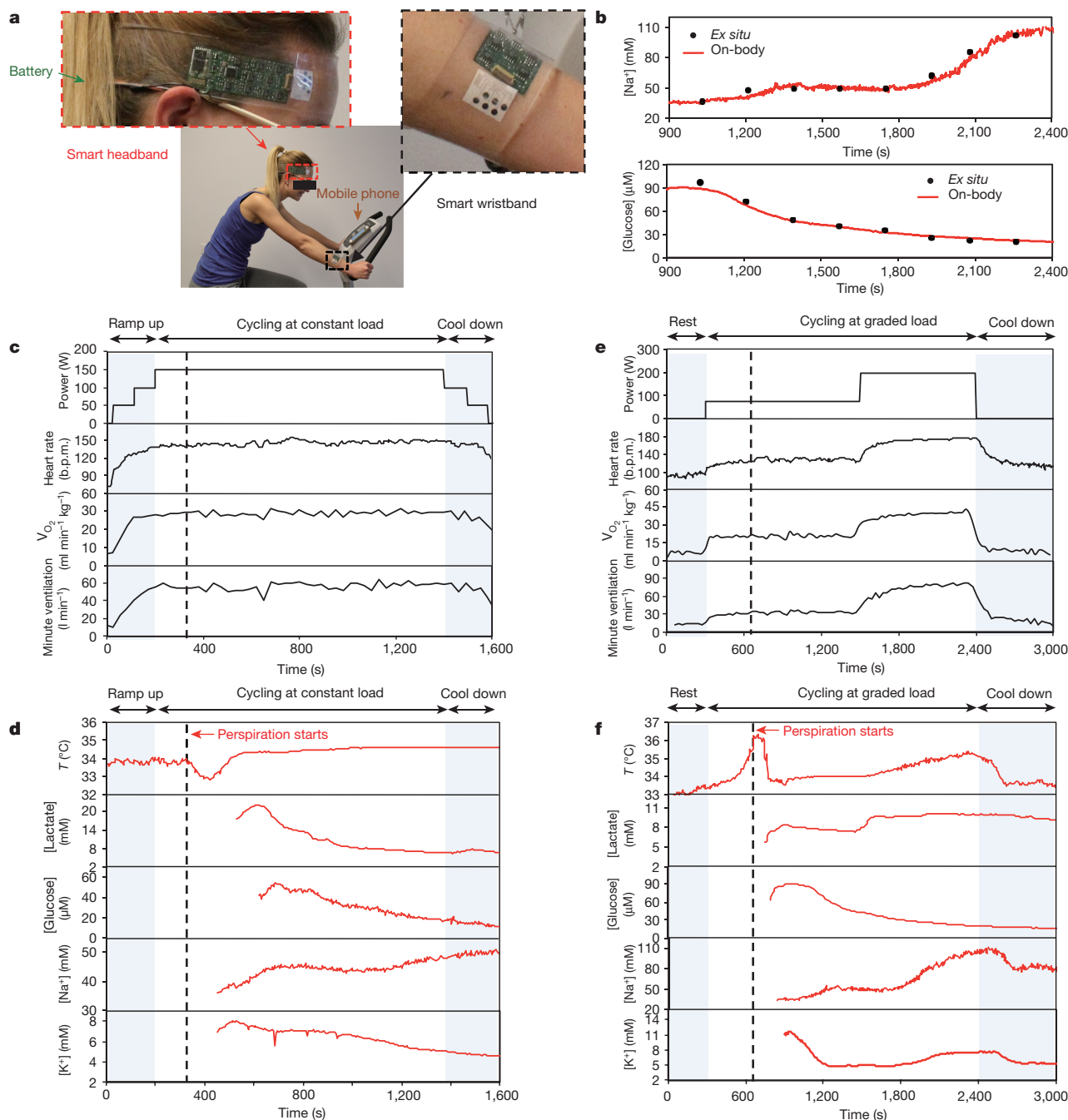


Figure 3 | On-body real-time perspiration analysis during stationary cycling. **a**, Photographs of a subject wearing a 'smart headband' and a 'smart wristband' during stationary cycling. **b**, Comparison of *ex situ* calibration data of the sodium and glucose sensors from the collected sweat samples with the on-body readings of the FISA during the stationary cycling exercise detailed in **f**. **c**, **d**, Constant-load exercise at 150 W: power output, heart rate (in beats per minute, b.p.m.), oxygen consumption (V_{O₂}) and pulmonary minute ventilation, as measured by external monitoring

involved a 5-min rest, 20-min cycling at 75 W followed by cycling at 200 W power output until volitional fatigue, and a 10-min recovery period (Fig. 3e and f). As demonstrated in Fig. 3e, the dramatic increase in the exercise power output from 75 W to 200 W immediately leads to abrupt elevations of heart rate, minute ventilation and V_{O₂}. Responses of the FISA during 75-W power output follow profiles similar to those observed during the constant-load study. After the power is raised, the sweat rate visibly increases, followed by a sharp increase in skin temperature and sweat [Na⁺] as well as a slight increase in [K⁺] (in three of the

systems (c) and the real-time sweat analysis results of the FISA worn on a subject's forehead (d). **e**, **f**, Graded-load exercise, involving a dramatic power increase from 75 W to 200 W: power output, heart rate, V_{O₂} and pulmonary minute ventilation, as measured by external monitoring systems (e) and the real-time analysis results using the FISA worn on a different subject's forehead (f). Data collection for each sensor took place when a sufficient sweat sample was present (see Methods).

seven subjects, [K⁺] remained stable). The relatively stable behaviour of [K⁺] is explained by its passive ion partitioning mechanism¹³. With the cessation of exercise, these physiological responses decrease and then remain stable. No apparent difference is observed for glucose concentration at different power output settings, a finding consistent with the response of blood glucose to graded, short-term exercise³⁰. The change in lactate concentration, on the other hand, varies between subjects. This observation can be attributed to the increase in both lactate excretion rate and sweat rate upon the increase of the workload³¹.

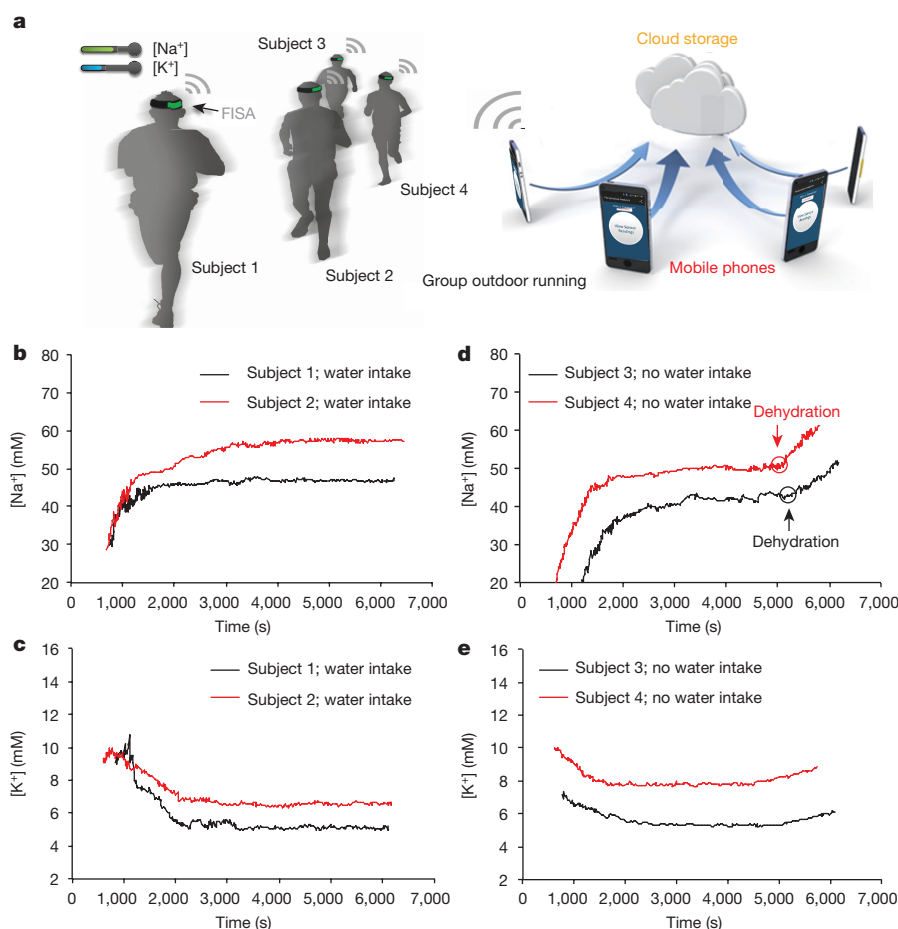


Figure 4 | Hydration status analysis during group outdoor running using the FISAs. **a**, Schematic illustration showing the group outdoor running trial based on wearable FISAs (packaged as 'smart headbands'). The data are transmitted to the user's cell phone and uploaded to cloud

Monitoring hydration status is of the utmost importance to athletes because fluid deficit impairs endurance performance and increases carbohydrate reliance³². To evaluate the utility of a FISA for effective and non-invasive identification of dehydration, real-time sweat $[\text{Na}^+]$ and $[\text{K}^+]$ measurements were conducted simultaneously on a group of subjects engaged in prolonged outdoor running trials (Fig. 4a). Figure 4b and c shows that sweat $[\text{Na}^+]$ and $[\text{K}^+]$ are stable throughout running in euhydration trials (with water intake of 150 ml per 5 min) after the initial $[\text{Na}^+]$ increase and $[\text{K}^+]$ decrease. On the other hand, a substantial increase in sweat $[\text{Na}^+]$ and a smaller increase in sweat $[\text{K}^+]$ (no clear increase in $[\text{K}^+]$ was observed in two out of six subjects) were observed in dehydration trials (without water intake) after 80 min when subjects had lost a large amount of water ($\sim 2.5\%$ of body weight) (Fig. 4d and e). *Ex situ* measurements of $[\text{Na}^+]$ and $[\text{K}^+]$ from collected sweat samples in Extended Data Fig. 10 also show similar phenomena. These trends are probably caused by increased blood serum $[\text{Na}^+]$ and $[\text{K}^+]$ with dehydration and increased neural stimulation, a conclusion in agreement with previous *ex situ* sweat analyses³³. Thus, sweat $[\text{Na}^+]$ can potentially serve as an important biomarker for monitoring dehydration. We believe that this wearable platform may enable new fundamental physiology studies through further on-body evaluation.

Thus, we have merged skin-conforming plastic-based sensors (five different sensors) and conventional commercially available integrated-circuit components (more than ten chips) at an unprecedented level of integration, not only to measure the output of an array of multiplexed and selective sensors, but also to obtain an accurate assessment via signal processing of the physiological state of the human subjects. This application could not have been realized by either of the technologies

servers. **b**, **c**, Representative real-time sweat sodium (**b**) and potassium (**c**) levels during an endurance run with water intake. **d**, **e**, Representative real-time sweat sodium (**d**) and potassium (**e**) levels during an endurance run without water intake.

(flexible sensors and silicon integrated circuits) alone, owing to their respective inherent limitations. The plastic-based device technologies lack the ability to implement sophisticated electronic functionalities for critical signal conditioning and processing. On the other hand, the silicon integrated-circuit technology does not provide sufficiently large active areas nor the intimate skin contact required to achieve stable and sensitive on-body measurements. Importantly, the entire system is mechanically flexible, thus delivering a practical wearable sensor technology that can be used for prolonged indoor and outdoor physical activities. This platform could be exploited or reconfigured for *in situ* analyses of other biomarkers within sweat and other human fluid samples to facilitate personalized and real-time physiological and clinical investigations. We envision that the large data sets that could be collected through such studies, along with voluntary community participation, would enable data-mining techniques with which to generate predictive algorithms for understanding the health status and clinical needs of individuals and society.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 June; accepted 23 November 2015.

- Kim, D. H. *et al.* Epidermal electronics. *Science* **333**, 838–843 (2011).
- Lipomi, D. J. *et al.* Skin-like pressure and strain sensors based on transparent elastic films of carbon nanotubes. *Nature Nanotechnol.* **6**, 788–792 (2011).
- Takei, K. *et al.* Nanowire active-matrix circuitry for low-voltage macroscale artificial skin. *Nature Mater.* **9**, 821–826 (2010).

4. Wu, W. *et al.* Piezoelectricity of single-atomic-layer MoS₂ for energy conversion and piezotronics. *Nature* **514**, 470–474 (2014).
5. McAlpine, M. C., Ahmad, H., Wang, D. & Heath, J. R. Highly ordered nanowire arrays on plastic substrates for ultrasensitive flexible chemical sensors. *Nature Mater.* **6**, 379–384 (2007).
6. Kaltenbrunner, M. *et al.* An ultra-lightweight design for imperceptible plastic electronics. *Nature* **499**, 458–463 (2013).
7. Xu, S. *et al.* Soft microfluidic assemblies of sensors, circuits, and radios for the skin. *Science* **344**, 70–74 (2014).
8. Wang, C. *et al.* User-interactive electronic skin for instantaneous pressure visualization. *Nature Mater.* **12**, 899–904 (2013).
9. Liang, J., Li, L., Niu, X., Yu, Z. & Pei, Q. Elastomeric polymer light-emitting devices and displays. *Nature Photon.* **7**, 817–824 (2013).
10. Kim, D. H., Ghaffari, R. & Rogers, J. A. Flexible and stretchable electronics for biointegrated devices. *Annu. Rev. Biomed. Eng.* **14**, 113–128 (2012).
11. Patel, S., Park, H., Bonato, P., Chan, L. & Rodgers, M. A review of wearable sensors and systems with application in rehabilitation. *J. Neuroeng. Rehabil.* **9**, 21 (2012).
12. Bandodkar, A. J. & Wang, J. Non-invasive wearable electrochemical sensors: a review. *Trends Biotechnol.* **32**, 363–371 (2014).
13. Sonner, Z. *et al.* The microfluidics of the eccrine sweat gland, including biomarker partitioning, transport, and biosensing implications. *Biomicrofluidics* **9**, 031301 (2015).
14. Schazmann, B. *et al.* A wearable electrochemical sensor for the real-time measurement of sweat sodium concentration. *Anal. Methods* **2**, 342–348 (2010).
15. Jia, W. *et al.* Electrochemical tattoo biosensors for real-time noninvasive lactate monitoring in human perspiration. *Anal. Chem.* **85**, 6553–6560 (2013).
16. Coyle, S. *et al.* BIOTEX—biosensing textiles for personalised healthcare management. *IEEE Trans. Inf. Technol. Biomed.* **14**, 364–370 (2010).
17. Bandodkar, A. J. *et al.* Epidermal tattoo potentiometric sodium sensors with wireless signal transduction for continuous non-invasive sweat monitoring. *Biosens. Bioelectron.* **54**, 603–609 (2014).
18. Rose, D. P. *et al.* Adhesive RFID sensor patch for monitoring of sweat electrolytes. *IEEE Trans. Biomed. Eng.* **62**, 1457–1465 (2015).
19. Boysen, T. C., Yanagawa, S., Sato, F. & Sato, K. A modified anaerobic method of sweat collection. *J. Appl. Physiol.* **56**, 1302–1307 (1984).
20. Speedy, D. B., Noakes, T. D. & Schneider, C. Exercise-associated hyponatremia: a review. *Emerg. Med.* **13**, 17–27 (2001).
21. Talar, M. S., Dewarrat, F., Huber, D. & Caduff, A. *In vivo* life sign application of dielectric spectroscopy and non-invasive glucose monitoring. *J. Non-Cryst. Solids* **353**, 4515–4517 (2007).
22. Derbyshire, P. J., Barr, H., Davis, F. & Higson, S. P. Lactate in human sweat: a critical review of research to the present day. *J. Physiol. Sci.* **62**, 429–440 (2012).
23. Webb, R. C. *et al.* Ultrathin conformal devices for precise and continuous thermal characterization of human skin. *Nature Mater.* **12**, 938–944 (2013).
24. Sprigle, S., Linden, M., McKenna, D., Davis, K. & Riordan, B. Clinical skin temperature measurement to predict incipient pressure ulcers. *Adv. Skin Wound Care* **14**, 133–137 (2001).
25. Zhu, J., Qin, Y. & Zhang, Y. Preparation of all solid-state potentiometric ion sensors with polymer-CNT composites. *Electrochem. Commun.* **11**, 1684–1687 (2009).
26. Jobst, G. *et al.* Thin-film microbiosensors for glucose-lactate monitoring. *Anal. Chem.* **68**, 3173–3179 (1996).
27. Nadel, E. R., Bullard, R. W. & Stolwijk, J. A. Importance of skin temperature in the regulation of sweating. *J. Appl. Physiol.* **31**, 80–87 (1971).
28. Cage, G. W., Wolfe, S. M., Thompson, R. H. & Gordon, R. S. Jr. Effects of water intake on composition of thermal sweat in normal human volunteers. *J. Appl. Physiol.* **29**, 687–690 (1970).
29. Patterson, M. J., Galloway, S. D. R. & Nimmo, M. A. Variations in regional sweat composition in normal human males. *Exp. Physiol.* **85**, 869–875 (2000).
30. Messonnier, L. A. *et al.* Lactate kinetics at the lactate threshold in trained and untrained men. *J. Appl. Physiol.* **114**, 1593–1602 (2013).
31. Buono, M. J., Lee, N. V. & Miller, P. W. The relationship between exercise intensity and the sweat lactate excretion rate. *J. Physiol. Sci.* **60**, 103–107 (2010).
32. Barr, S. I. Effects of dehydration on exercise performance. *Can. J. Appl. Physiol.* **24**, 164–172 (1999).
33. Morgan, R. M., Patterson, M. J. & Nimmo, M. A. Acute effects of dehydration on sweat composition in men during prolonged exercise in the heat. *Acta Physiol. Scand.* **182**, 37–43 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements The sensor design, characterization and testing aspects of this work were supported by the Berkeley Sensor and Actuator Center, and National Institutes of Health grant number P01 HG000205. The sensor fabrication was performed in the Electronic Materials (E-MAT) laboratory funded by the Director, Office of Science, Office of Basic Energy Sciences, Material Sciences and Engineering Division of the US Department of Energy under contract number DE-AC02-05CH11231. K.C. acknowledges funding from the NSF Nanomanufacturing Systems for mobile Computing and Energy Technologies (NASCENT) Center. H.O. acknowledges support from a Japan Society for the Promotion of Science (JSPS) Fellowship. We thank J. Bullock, C. M. Sutter-Fella, H. W. W. Nyein, Z. Shahpar, M. Zhou, E. Wu and W. Chen for their help.

Author Contributions W.G., S.E. and A.J. conceived the idea and designed the experiments. W.G., S.E., H.Y.Y.N. and S.C. led the experiments (with assistance from K.C., A.P., H.M.F., H.O., H.S., H.O., D.K., D.-H.L.). W.G., S.E., A.P., G.A.B., R.W.D. and A.J. contributed to data analysis and interpretation. W.G., S.E., H.Y.Y.N., G.A.B. and A.J. wrote the paper and all authors provided feedback.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.J. (ajavey@eecs.berkeley.edu).

METHODS

Materials. Selectophore grade sodium ionophore X, bis(2-ethylexyl) sebacate (DOS), sodium tetrakis[3,5-bis(trifluoromethyl)phenyl] borate (Na-TFPB), high-molecular-weight polyvinyl chloride (PVC), tetrahydrofuran, valinomycin (potassium ionophore), sodium tetraphenylborate (NaTPB), cyclohexanone, polyvinyl butyral resin BUTVAR B-98 (PVB), sodium chloride (NaCl), 3,4-ethylenedioxythiophene (EDOT), poly(sodium 4-styrenesulfonate) (NaPSS), glucose oxidase (from *Aspergillus niger*), chitosan, single-walled carbon nanotubes, iron (III) chloride, potassium ferricyanide (III), multiwall carbon nanotubes and block polymer PEO-PPO-PEO (F127) were obtained from Sigma Aldrich. L-lactate oxidase (>80 activity units per milligram) was procured from Toyobo Corp. and PBS (pH 7.2) was obtained from Life Technologies. Moisture-resistant 100- μm -thick PET was purchased from McMaster-Carr.

Fabrication of electrode arrays. The fabrication process of the electrode arrays is detailed in Extended Data Fig. 1. Briefly, the sensor arrays on PET were patterned by photolithography using positive photoresist (Shipley Microposit S1818) followed by 30 nm Cr/50 nm Au deposited via electron-beam evaporation and lift-off in acetone. A 500-nm parylene C insulation layer was then deposited in a SCS Labcoter 2 Parylene Deposition System. Subsequently, photolithography was used to define the final electrode area (3 mm diameter) followed by O_2 plasma etching for 450 s at 300 W to remove the parylene completely. Electron-beam evaporation was then performed to pattern 180-nm Ag onto the electrode areas, followed by lift-off in acetone. The Ag patterns on working electrode area were dissolved in a 6-M HNO_3 solution for 1 min. The Ag/AgCl reference electrodes were obtained by injecting 10 μl 0.1-M FeCl_3 solution on top of each Ag reference electrode using a micropipette for 1 min.

Design of electrochemical sensors. For amperometric glucose and lactate sensors, a two-electrode system where Ag/AgCl acts as both reference and counter electrode was chosen to simplify circuit design and to facilitate system integration. The two-electrode system is a common strategy for low-current electrochemical sensing^{34,35}. The output currents (between the working electrode and the Ag/AgCl reference/counter electrode) of the glucose and lactate sensors could be converted to a voltage potential through a transimpedance amplifier. It is known that amperometric sensors with larger area provide larger current signal. Considering the low concentration of glucose in sweat, we designed the sensors to be 3 mm in diameter to obtain a high current.

Preparation of Na^+ and K^+ selective sensors. The Na^+ selective membrane cocktail consisted of Na ionophore X (1% weight by weight, w/w), Na-TFPB (0.55% w/w), PVC (33% w/w), and DOS (65.45% w/w). 100 mg of the membrane cocktail was dissolved in 660 μl of tetrahydrofuran¹⁷. The K^+ -selective membrane cocktail was composed of valinomycin (2% w/w), NaTPB (0.5%), PVC (32.7% w/w), and DOS (64.7% w/w). 100 mg of the membrane cocktail was dissolved in 350 μl of cyclohexanone. The ion-selective solutions were sealed and stored at 4 °C. The solution for the PVB reference electrode was prepared by dissolving 79.1 mg PVB and 50 mg of NaCl into 1 ml methanol³⁶. 2 mg F127 and 0.2 mg of multiwall carbon nanotubes were added into the reference solution to minimize the potential drift²⁵.

Poly(3,4-ethylenedioxythiophene) PEDOT:PSS was chosen as the ion-electron transducer to minimize the potential drift of the ISEs³⁷ and deposited onto the working electrodes by galvanostatic electrochemical polymerization with an external Ag/AgCl reference electrode from a solution containing 0.01-M EDOT and 0.1-M NaPSS. A constant current of 14 μA (2 mA cm^{-2}) was applied to produce polymerization charges of 10 mC onto each electrode.

Ion-selective membranes were then prepared by drop-casting 10 μl of the Na^+ -selective membrane cocktail and 4 μl of the K^+ -selective membrane cocktail onto their corresponding electrodes. The common reference electrode for the Na^+ and K^+ ISEs was modified by casting 10 μl of reference solution onto the Ag/AgCl electrode. The modified electrodes were left to dry overnight. The sensors could be used without pre-conditioning (with a small drift of $\sim 2\text{--}3\text{ mV h}^{-1}$). However, to obtain the best performance for long-term continuous measurements such as dehydration studies, the ion-selective sensors were covered with a solution containing 0.1-M NaCl and 0.01-M KCl through microinjection (without contact to glucose and lactate sensors) for 1 h before measurements. This conditioning process was important to minimize the potential drift further.

Preparation of lactate and glucose sensors. 1% chitosan solution was first prepared by dissolving chitosan in 2% acetic acid and magnetic stirring for about 1 h; next, the chitosan solution was mixed with single-walled carbon nanotubes (2 mg ml^{-1}) by ultrasonic agitation over 30 min to prepare a viscous solution of chitosan and carbon nanotubes. To prepare the glucose sensors, the chitosan/carbon nanotube solution was mixed thoroughly with glucose oxidase solution (10 mg ml^{-1} in PBS of pH 7.2) in the ratio 2:1 (volume by volume). A Prussian blue mediator layer was deposited onto the Au electrodes by cyclic voltammetry from 0 V to 0.5 V (versus Ag/AgCl) for one cycle at a scan rate of 20 mV s^{-1} in a fresh solution containing 2.5 mM FeCl_3 , 100 mM KCl, 2.5 mM $\text{K}_3\text{Fe}(\text{CN})_6$, and 100 mM HCl. A thinner Prussian

blue layer can provide better sensitivity, which is essential for low-glucose-level measurements in sweat. The glucose sensor was obtained by drop-casting 3 μl of the glucose oxidase/chitosan/carbon nanotube solution onto the Prussian blue/Au electrode. For the lactate sensors, the Prussian blue mediator layer was deposited onto the Au electrodes by cyclic voltammetry from -0.5 V to 0.6 V (versus Ag/AgCl) for 10 cycles at 50 mV s^{-1} in a fresh solution containing 2.5 mM FeCl_3 , 100 mM KCl, 2.5 mM $\text{K}_3\text{Fe}(\text{CN})_6$, and 100 mM HCl. A thicker Prussian blue layer can provide a wider linear response range, which is crucial for lactate measurement in sweat. 3 μl of the chitosan/carbon nanotube solution was drop-cast onto the Prussian blue/Au electrode and dried in the ambient environment; the electrode was later covered with 2 μl of lactate oxidase solution (40 mg ml^{-1}) and finally 3 μl of the chitosan/carbon nanotube solution. The sensor arrays were allowed to dry overnight at 4 °C with no light. The solutions were stored at 4 °C when not in use.

Signal conditioning, processing and wireless transmission circuit design. The circuit diagram of the analogue signal-conditioning block of the FISA is shown in Extended Data Fig. 4. At the core of our system we used an ATmega328P (Atmel 8-bit) microcontroller that could be programmed on-board through an in-circuit serial programming interface. This microcontroller is compatible with the popular Arduino development environment, and is commonly used in autonomous systems with low power and low cost requirements. By exploiting the microcontroller's built-in 10-bit analogue-to-digital converter block as well as its computational and serial communication capability, we relayed the signals (as transduced by our sensor module and as conditioned by our analogue circuitry) to the Bluetooth transceiver.

The conditioning path for each sensor was implemented in relation to the corresponding sensing mode. In the case of the amperometric-based glucose and lactate sensors, the originally generated signal was in the form of electrical current. Therefore, in the respective signal conditioning paths, we first used a transimpedance amplifier stage to convert the signal current into voltage. In our electrical current measurements, the direction of the current was from the shared Ag/AgCl reference/counter electrode towards the working electrode of each of the glucose and lactate sensors, which would result in a negative transimpedance output voltage. Hence, for both glucose and lactate paths, the transimpedance amplifiers were followed by inverter stages to make the respective voltage signals positive, since the analogue-to-digital converter stage took only positive input values. The feedback resistors in each of the transimpedance sections was chosen (1 M Ω for the glucose path and 0.5 M Ω for the lactate path) such that the converted voltage signal could be finely resolved, while staying within the input voltage range of the analogue-to-digital converter stage of the microcontroller. The current sensing signal paths were capable of measuring current levels as low as 1 nA, which was much lower than the minimum signal in our measurements (tens of nanoamperes). In this implementation, with the transimpedance amplifier at the front-end, the Ag/AgCl reference/counter electrode of the amperometric-based sensors needed to be grounded. This requirement prevented us from grounding the shared PVB reference electrodes in the potentiometric-based sensors, because the potential difference between the Ag/AgCl reference and PVB electrodes changes in the presence of different chloride ion concentrations (Extended Data Fig. 2b). In the case of the ISE-based sensors, the generated signals were essentially the voltage differences between the PVB-coated shared reference electrode and the working electrode of the respective sensors. Therefore, without grounding the PVB electrode, we measured the difference in potential of the floating ISE working and shared electrodes directly. To this end, the signal conditioning paths of the potentiometric-based sensors included a voltage buffer interfacing the respective working and reference electrodes, followed by a differential amplifier to effectively implement an instrumentation amplifier configuration. With this approach we ensured that the voltage-sensing and current-sensing paths were electrically isolated. Furthermore, the differential sensing stage also helped to minimize the unwanted common-mode interferences which would have otherwise degraded the fidelity of our sensor readings. Also, the high impedance nature of the ISE-based sensors³⁷ required the use of high-impedance voltage buffers to ensure accurate open voltage measurement as intended.

All the analogue signal conditioning paths concluded with a corresponding unity gain four-pole low pass filter, each with a -3 dB frequency at 1 Hz to minimize the noise and interference in our measurements. The choice of using active filters in our system also gave us flexibility in tuning the gain in our signal-conditioning path if needed. The low pass filters were connected to the analogue-to-digital converter stage of the microcontroller, to facilitate the conversion of the filtered analogue signals to their respective digital forms. In our implementation, each of the analogue signal conditioning paths were electrically characterized to validate the linear output response of the channels with respect to the corresponding electrical input signals mimicking the sensor output signals. For this characterization step, electrical current was applied as an input to the glucose and lactate channel terminals to model the respective amperometric-based sensor output and differential voltage was applied at the terminals of the sodium

and potassium channels to model the corresponding potentiometric-based sensor output. As illustrated in Extended Data Fig. 5a–d, all four signal-conditioning channels demonstrated an excellent linear response (correlation factor $R^2 = 1$). To eliminate the non-ideal effects such as voltage offset and to obtain precise signal readings, the exact numerical linear relationship between output and input was obtained to map the original input signal to the analogue circuit readouts, which in turn allowed for subsequent signal calibration and processing at the software level. Upon processing and averaging the data, the microcontroller was exploited to relay the data to the Bluetooth module for wireless transmission.

Power delivery to the FISA. The FISA was powered by a single rechargeable lithium-ion polymer battery with a nominal voltage of 3.7 V of a desired capacity (a representative 105-mAh battery is illustrated in Extended Data Fig. 5e and f). The protection circuitry included protects the battery against unwanted output shorts and over-charging. Step-up direct current/direct current converters were used to produce a fixed, regulated output of +5 V for the microcontroller and +3.3 V for the Bluetooth modules. This regulated output also served as the positive power supply for the analogue peripheral components. The negative power supply (–5 V) for the analogue peripheral components was implemented through the use of inverting charge pump direct current/direct current converters that produce negative regulated outputs.

The custom mobile application design. A mobile application (the Perspiration Analysis App) was designed to accompany the FISA and to provide a user-friendly interface for data display and aggregation (Extended Data Fig. 3). To use this application, first, the user should wear the FISA and open the Perspiration Analysis App on the mobile device. The application establishes a secure Bluetooth connection to the FISA. Subsequently, it receives and displays the stream of data that are transmitted in real time from the FISA. The application is capable of plotting a graph of these values versus time during the user's physical activities. The data and graphs can be stored on the device, uploaded to cloud servers online, and can be shared via social media. Additionally, the application keeps track of the duration of exercise as well as the distance travelled. Although the current implementation was programmed in the Android environment, similar application interfaces could easily be developed in other popular mobile operating systems such as iOS.

The characterization of the sensors. A set of electrochemical sensors was characterized to explore their reproducibility in solutions of target analytes. Extended Data Fig. 6a–d show that Na^+ and K^+ sensors had a relative standard deviation of $\sim 1\%$ in sensitivity while glucose and lactate sensors had a relative standard deviation of $\sim 5\%$ in sensitivity. However, there are differences in absolute potential values for ISEs in the same solution. Therefore, one-point calibration in a standard solution containing 1 mM KCl and 10 mM NaCl was performed for Na^+ and K^+ sensors before each use. The measured potential of ISEs in the standard solution was then set to zero by the microcontroller. (Such calibration is similar to what is done in commercial finger-stick glucose sensors.) No calibration was needed for the glucose and lactate sensors. Long-term stability of the sensors was also evaluated over a period of four weeks using five different sensor arrays each week (Extended Data Fig. 6e–h). It was observed that the Na^+ and K^+ sensors had approximately the same sensitivities of 62.5 mV and 59.5 mV per decade of concentration, respectively, in ambient conditions. The sensitivities of the glucose and lactate sensors were similarly maintained within 5% of their original values over the four-week period when stored at 4°C. The glucose and lactate sensors were characterized chronoamperometrically using a Gamry Electrochemical Potentiostat (Fig. 2a and b). Owing to Faraday and capacitive currents³⁸, the responses of both sensors showed drift initially but stabilized within 1 min of the data recording. The *in vitro* temperature compensation experiments (Fig. 2h) were performed continuously using the same sensor in four Petri dishes containing solutions at different temperatures on different hot plates. The convection and non-uniform distributions of solution temperature could result in noticeable noise in the signal measurements.

For continuous use, all the sensors displayed excellent stability over the entire exercise period. The sensor array could be repeatedly used for continuous temperature and sweat electrolyte monitoring. However, the glucose and lactate responses degraded beyond the exercise period (after two hours) owing to decreased enzyme activity. The sensor–FPCB interface allows for convenient replacement of the fresh sensor arrays for subsequent use.

Analysis of the effect of mechanical deformation on the sensors was performed by repeatedly bending the Na^+ , glucose sensors, and temperature sensors (radius of curvature, 1.5 cm) as well as the FPCB (radius of curvature, 3 cm) for a total of 60 cycles (Extended Data Fig. 8). Performance of the sensors was recorded after every 30 cycles. Continuous measurement on sensor performance during bending and no bending was also performed.

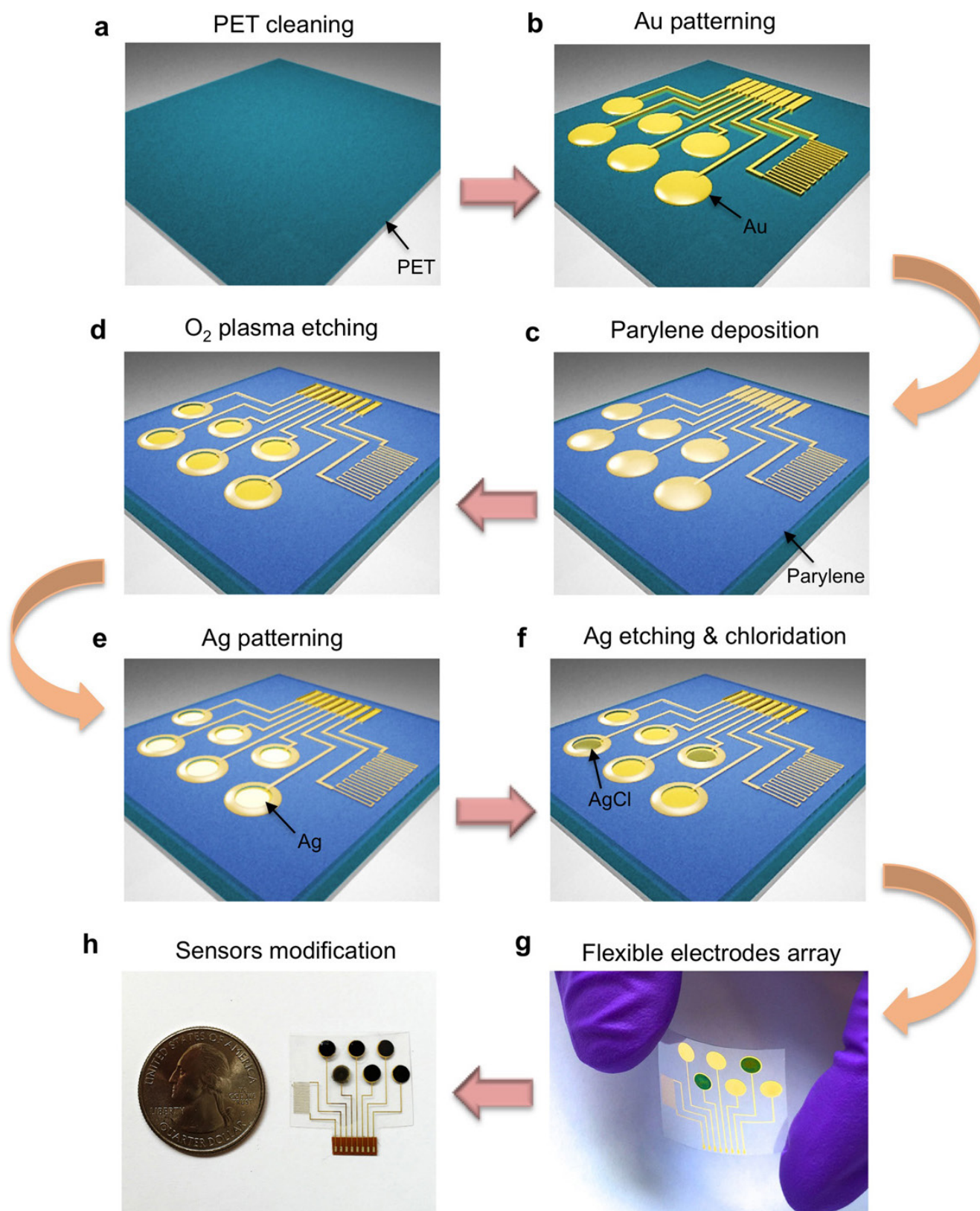
Ex situ evaluation of the sweat samples. *Ex situ* sensor performance was also conducted by testing sweat samples collected from the subjects' foreheads. Sweat samples were collected every 2–4 min by scratching their foreheads with microtubes, and subjects' foreheads were wiped and cleaned with gauze after every sweat collection¹⁹. The changes of $[\text{Na}^+]$ and $[\text{K}^+]$ during euhydration and dehydration trials were also studied *ex situ* in the same manner. The calibration of the

sensor arrays was performed before *ex situ* measurements using artificial sweat containing 22 mM urea, 5.5 mM lactic acid, 3 mM NH_4^+ , 0.4 mM Ca^{2+} , 50 μM Mg^{2+} and 25 μM uric acid with varying glucose concentrations of 0–200 μM , $[\text{K}^+]$ of 1–16 mM and $[\text{Na}^+]$ of 10–160 mM.

The setup of FISA for on-body testing. A water-absorbent thin rayon pad was placed between the skin and the sensor array during on-body experiments to absorb and maintain sufficient sweat for stable and reliable sensor readings, and to prevent direct mechanical contact between the sensors and skin. The pad could absorb about 10 μl of sweat, which was sufficient to provide stable sensor readings. During on-body tests, the newly generated sweat would refill the pad and 'rinse away' the old sweat. The on-body measurement results were also consistent with *ex situ* tests using freshly collected sweat samples. Assuming a single-centred flow model¹³, the best-case sampling interval can be calculated to be less than 1 min, based on the sweat rate ($\sim 3\text{--}4\text{ mg min}^{-1}\text{ cm}^{-2}$)³⁹ and the pad size ($1.5\text{ cm} \times 2\text{ cm} \times 50\text{ }\mu\text{m}$). The intrinsic response time of FISA was smaller than the body's response time to the changes in physiological conditions. An increase in temperature was observed when the 'smart headband' or 'smart wristband' was worn owing to the use of the plastic substrate on skin. Although this may result in a small error in measuring the actual skin temperature, it should be noted that this does not have an impact on the measurement of the electrolytes and metabolites, owing to the on-board temperature calibration. To ensure the fidelity of sensor readings further, the data collection of each channel took place when a sufficient sweat sample was present, as shown by the stabilization of the sensor readings (varying within 10% of the readings of the continuous five data points) within the physiologically relevant range: $[\text{Na}^+]$, 20–120 mM; $[\text{K}^+]$, 2–16 mM; glucose concentration, 0–200 μM ; and lactate concentration, 2–30 mM.

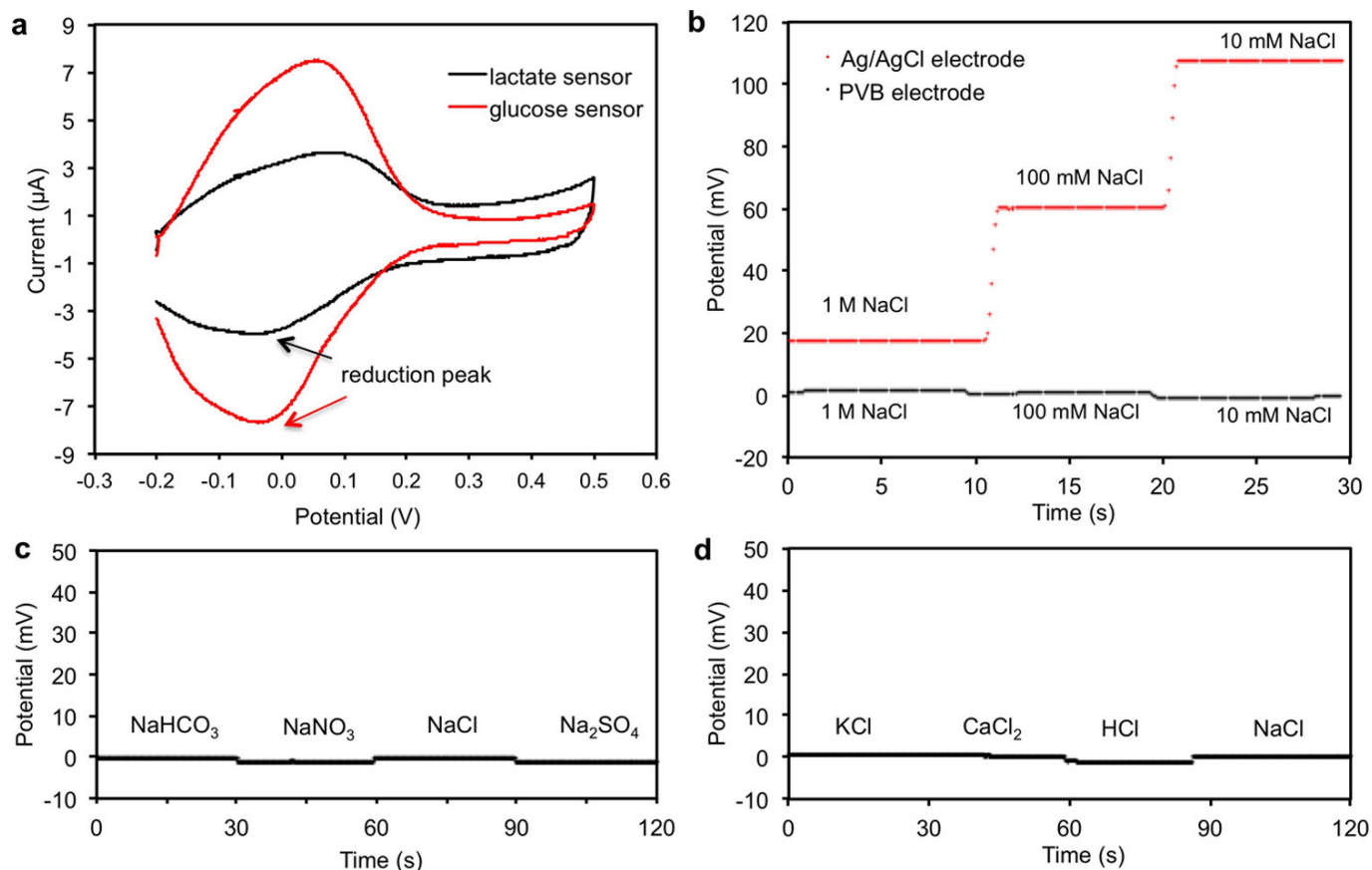
On-body sweat analysis. The on-body evaluation of the FISA was performed in compliance with the protocol that was approved by the institutional review board at the University of California, Berkeley (CPHS 2014-08-6636). 26 healthy subjects (4 females and 22 males), aged 20–40, were recruited from the University of California, Berkeley campus and the neighbouring community through advertisement by posted notices, word of mouth, and email distribution. All subjects gave written, informed consent before participation in the study. The study was conducted as three trials: constant workload cycle ergometry, graded workload cycle ergometry, and outdoor running. Constant workload cycle ergometry was conducted on 14 volunteers (4 females and 10 males between the ages of 20 and 40). The graded cycle ergometry was conducted on 7 male volunteers (who were also involved in the constant workload cycle study). 12 male volunteers between the ages of 20 and 40 were recruited for the outdoor running study. An electronically braked leg-cycle ergometer (Monark Ergonomic 839E, Monark Exercise AB) was used for cycling trials, which included real-time monitoring of heart rate, oxygen consumption (V_{O_2}), and pulmonary minute ventilation. The power output was calibrated and monitored through the ergometer. Heart rate was measured using a Tickr heart rate monitor (Wahoo fitness), and V_{O_2} and minute ventilation were continuously recorded throughout trials via an open-circuit, automated, indirect calorimetry system (TrueOne metabolic system; ParvoMedics). The FISAs were packaged inside traditional sweatbands during the indoor and outdoor trials. The sensor arrays were calibrated, and the subjects' foreheads and wrists were cleaned with alcohol swabs and gauze before sensors were worn on-body. For the constant workload cycling trial subjects were cycling at 50 W with 50-W increments every 90 s up to 150 W, and 20 min of cycling at 150 W. The power output was then decreased by 50 W every 90 s. The graded workload trial consisted of 5 min of seated rest followed by cycling at 75 W for 20 min and then cycling at 200 W until fatigue followed by a 10-min rest. The outdoor running trial was conducted with a group of 12 subjects in which 6 were instructed to drink 150 ml water every 5 min and 6 did not drink water throughout the trial. Subjects consented to run until volitional fatigue at a self-selected pace ($8\text{--}12\text{ km h}^{-1}$) and the Na^+ and K^+ sensor responses (from their foreheads) were recorded.

34. Kudo, H. *et al.* A flexible and wearable glucose sensor based on functional polymers with Soft-MEMS techniques. *Biosens. Bioelectron.* **22**, 558–562 (2006).
35. Bandodkar, A. J. *et al.* Tattoo-based noninvasive glucose monitoring: a proof-of-concept study. *Anal. Chem.* **87**, 394–398 (2015).
36. Guinovart, T., Crespo, G. A., Rius, F. X. & Andrade, F. J. A reference electrode based on polyvinyl butyral (PVB) polymer for decentralized chemical measurements. *Anal. Chim. Acta* **821**, 72–80 (2014).
37. Bobacka, J. Potential stability of all-solid-state ion-selective electrodes using conducting polymers as ion-to-electron transducers. *Anal. Chem.* **71**, 4932–4937 (1999).
38. Grieshaber, D., MacKenzie, R., Vörös, J. & Reimhult, E. Electrochemical biosensors—sensor principles and architectures. *Sensors* **8**, 1400–1458 (2008).
39. Vimeiro-Gomes, A. C. Comparison of sweat rate during graded exercise and the local rate induced by pilocarpine. *Braz. J. Med. Biol. Res.* **38**, 1133–1139 (2005).



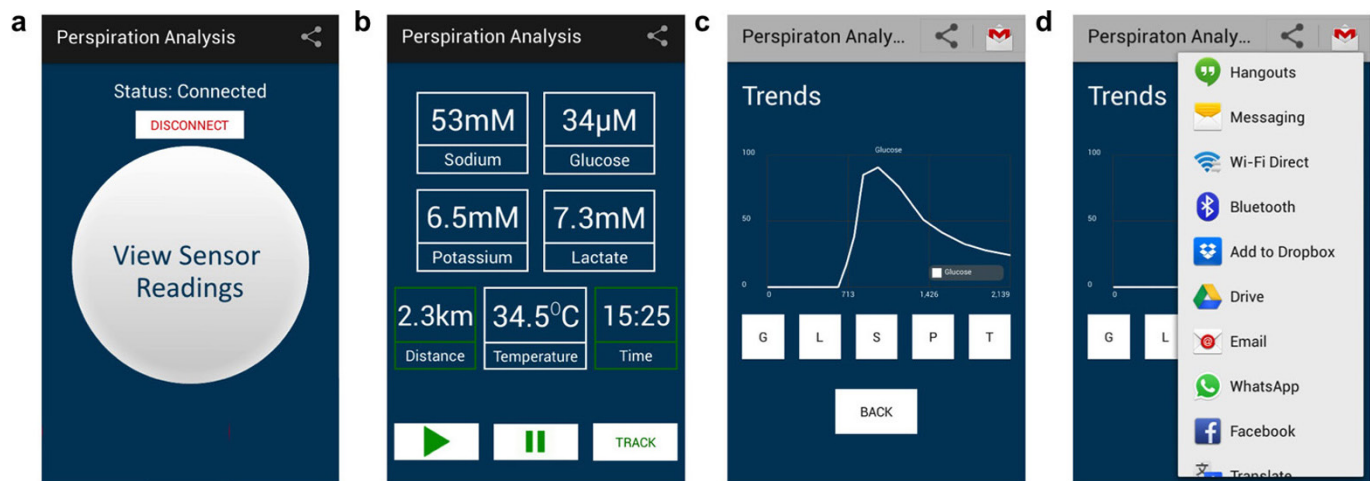
Extended Data Figure 1 | Fabrication process of the flexible sensor array. **a**, PET cleaning using acetone, isopropanol and O₂ plasma etching. **b**, Patterning of Cr/Au electrodes using photolithography, electron-beam evaporation and lift-off in acetone. **c**, Parylene insulating layer deposition. **d**, Photolithography and O₂ plasma etching of parylene in

the electrode areas. **e**, Electron-beam deposition of the Ag layer followed by lift-off in acetone. **f**, Ag etching on the Au working electrode area and Ag chloridation on the reference electrode area. **g**, Optical image of the flexible electrode array. **h**, Photograph of the multiplexed sensor array after surface modification.

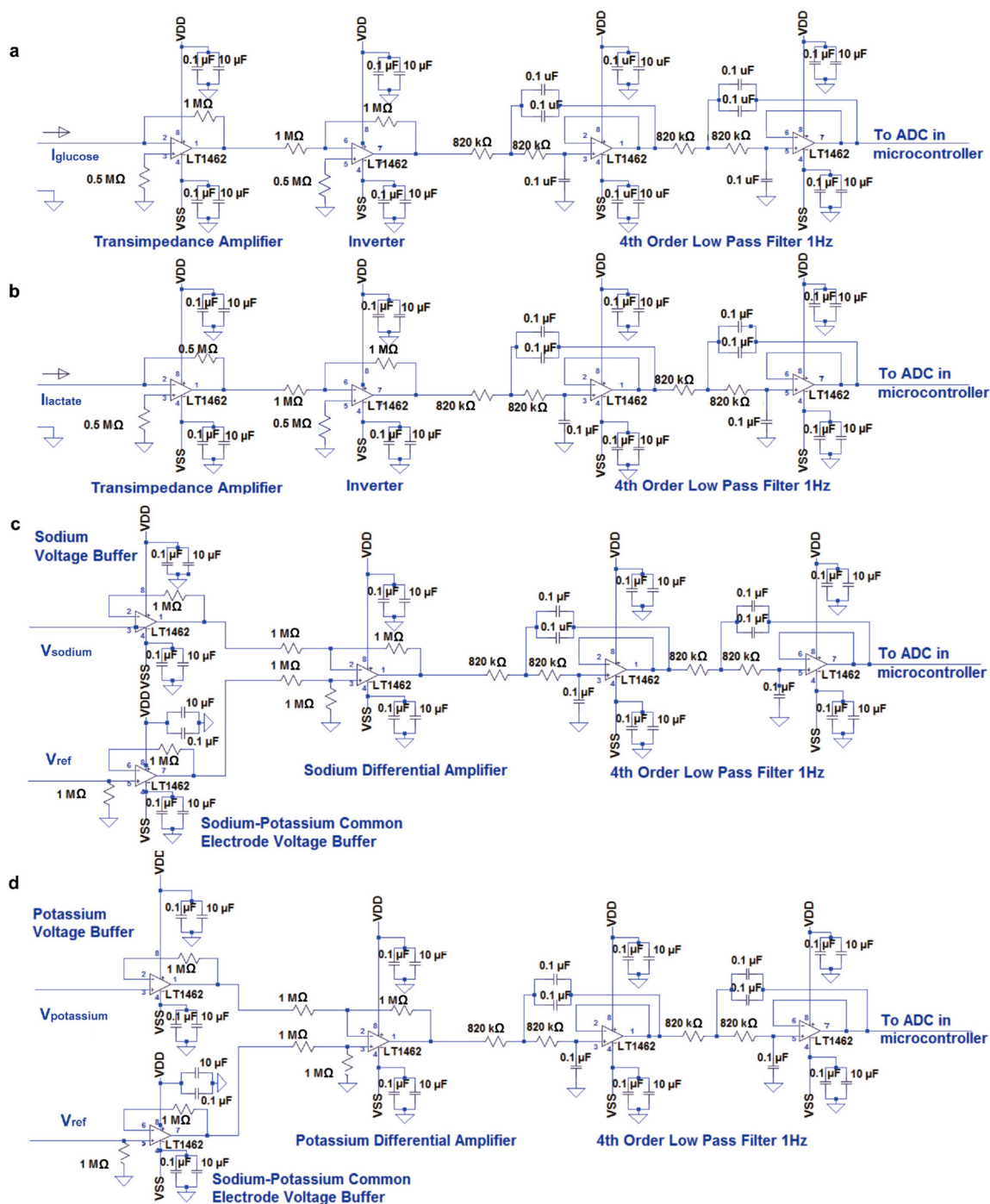


Extended Data Figure 2 | The characterizations of the modified electrodes. **a**, Cyclic voltammetry of the amperometric glucose and lactate sensors using Prussian blue as a mediator in PBS (pH 7.2). Scan range, -0.2 V to 0.5 V; scan rate, 50 mV s^{-1} . **b**, Potential stability of a PVB-coated Ag/AgCl electrode and a solid-state Ag/AgCl reference electrode (versus

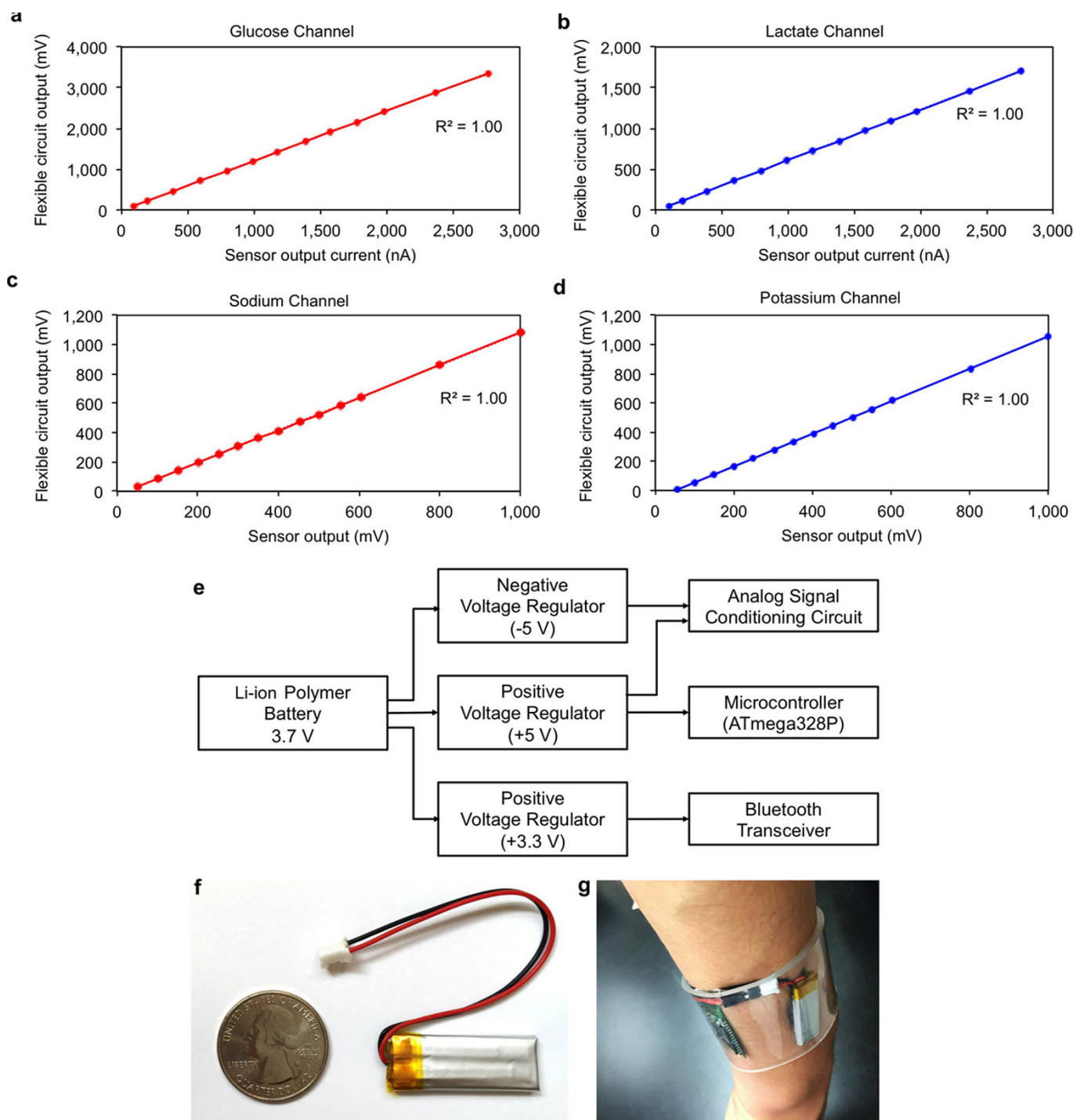
commercial aqueous Ag/AgCl electrode) in different NaCl solutions. **c**, **d**, The stability of a PVB-coated reference electrode in solutions containing 50 mM NaCl and 10 mM of different anionic (**c**) and cationic (**d**) solutions. Data recording was paused for 30 s for each solution change in **b–d**.



Extended Data Figure 3 | The custom-developed mobile application for data display and aggregation. a, The home page of the application after Bluetooth pairing. **b,** Real-time data display of sweat analyte levels as well as skin temperature during exercise. **c,** Real-time data progression of individual sensor. **d,** Available data sharing and uploading options.

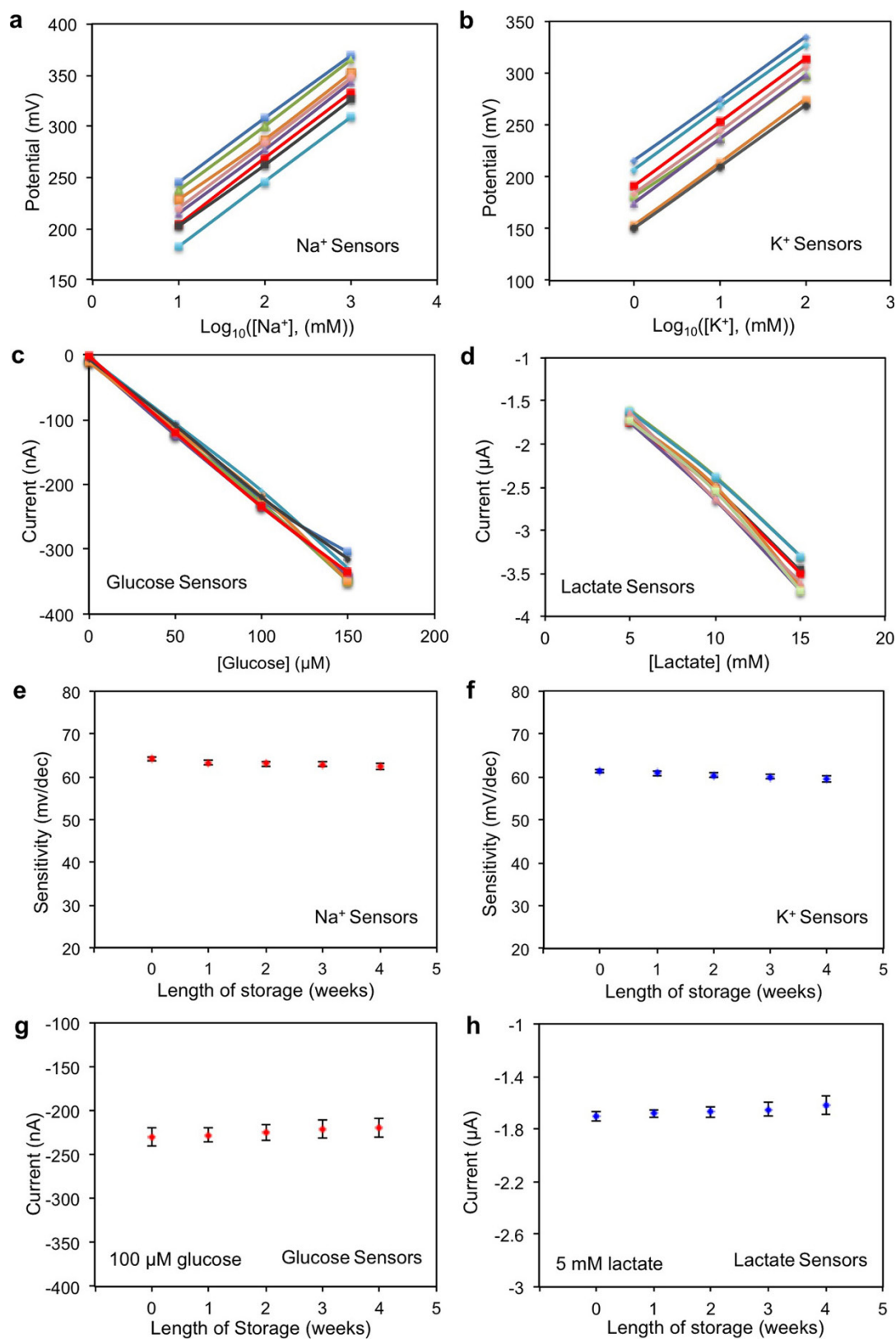


Extended Data Figure 4 | Schematic diagram of signal-conditioning circuit. a–d, Signal conditioning circuits for (a) glucose, (b) lactate, (c) sodium and (d) potassium channels. VDD and VSS represent the positive and negative power supplies, respectively. LT1462 is the integrated-circuit chip part.



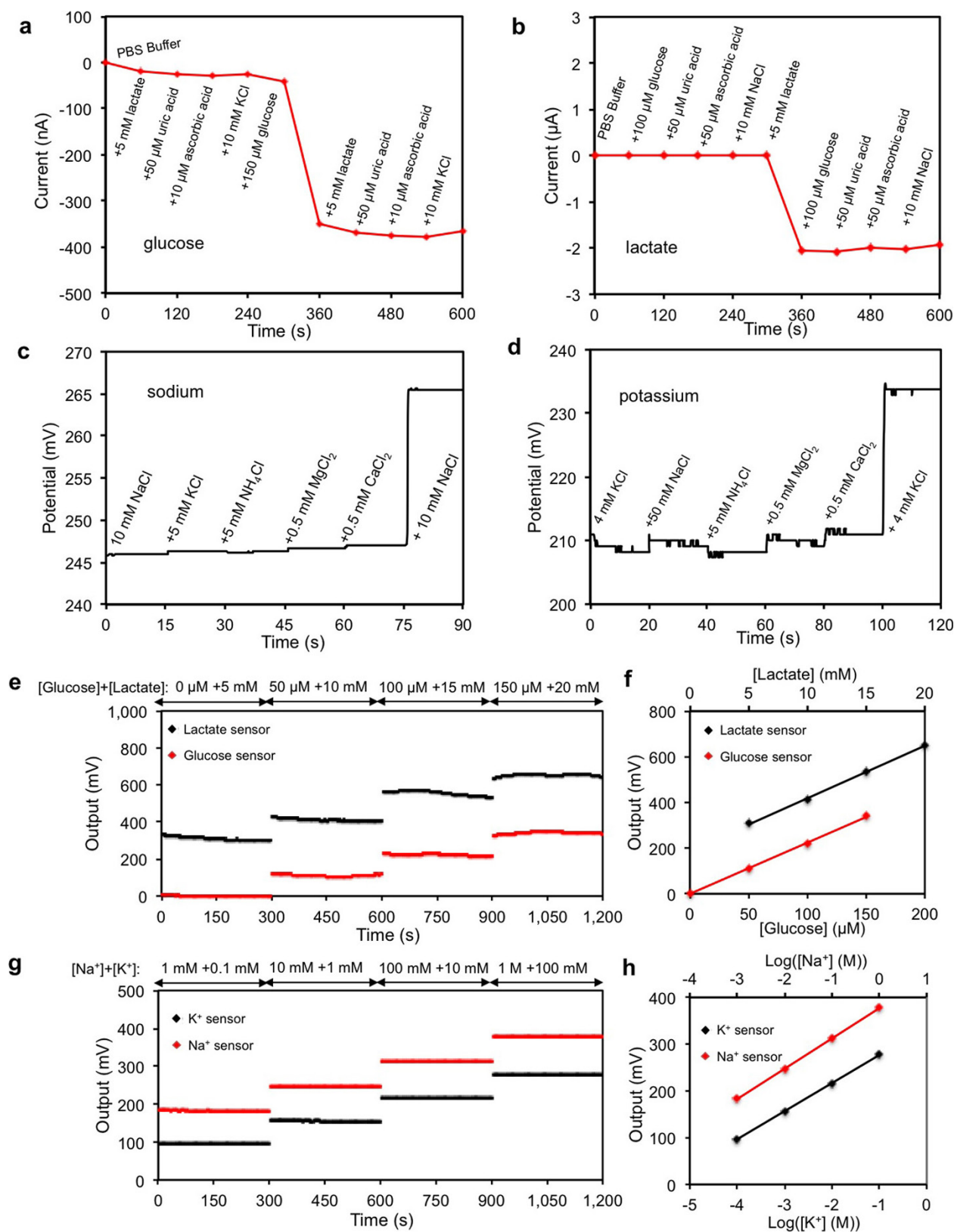
Extended Data Figure 5 | The calibration and power delivery of the FISA. **a–d**, Flexible PCB calibration for glucose (**a**), lactate (**b**), sodium (**c**) and potassium (**d**) channels. **e**, Power delivery diagram of the system. **f**, Photograph of a small rechargeable battery module used in

the current work (placed next to a quarter-dollar coin for comparison). **g**, Representative photograph of the power delivery package inside a transparent wristband on a subject's wrist.



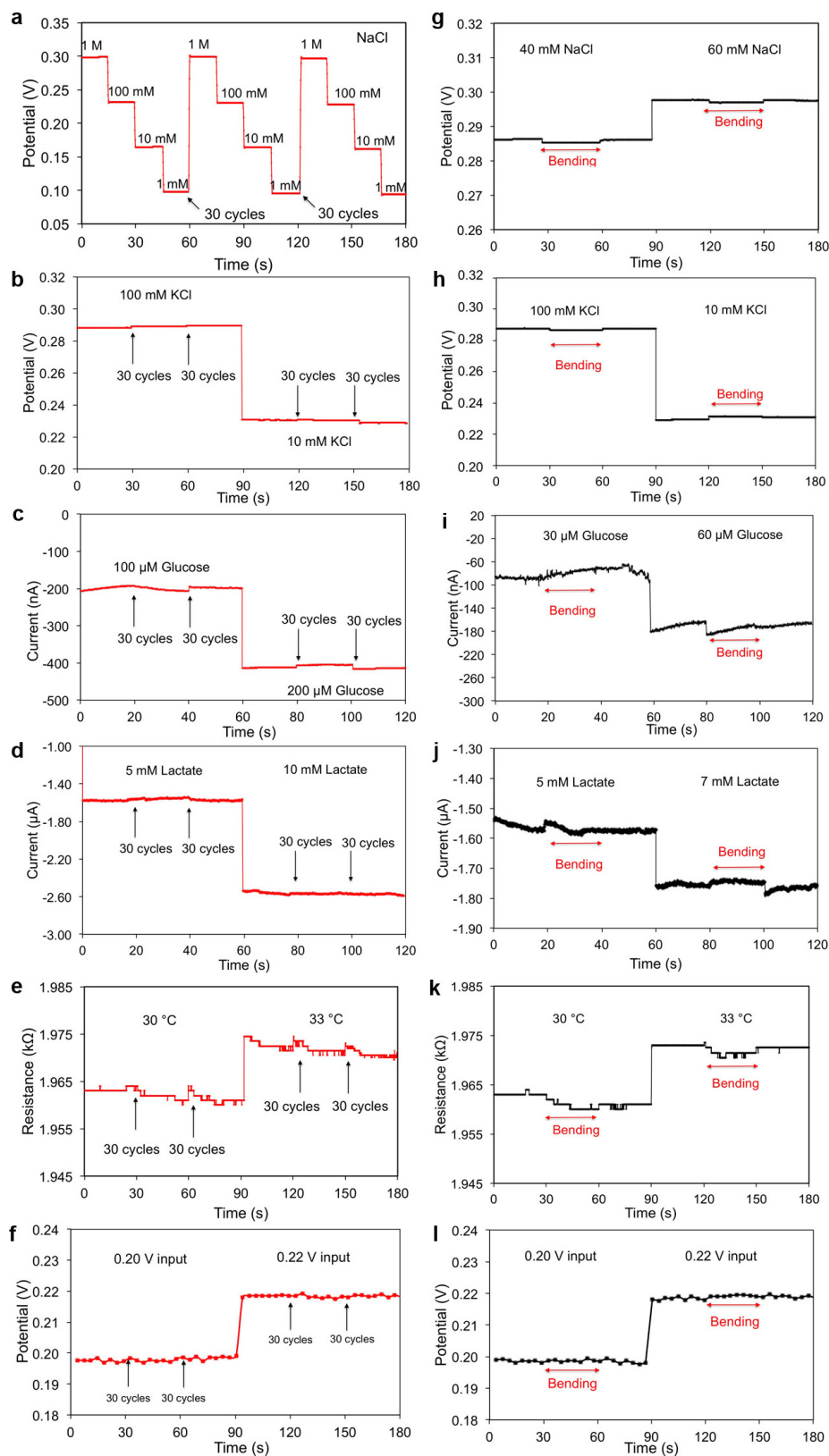
Extended Data Figure 6 | Reproducibility and long-term stability of the biosensors. **a–d**, The reproducibility of the sodium (**a**), potassium (**b**), glucose (**c**) and lactate (**d**) sensors (eight samples for each kind of sensor). **e–h**, The long-term stability of the sodium (**e**), potassium (**f**),

glucose (**g**) and lactate (**h**) sensors. Sensitivity is measured in millivolts per decade of concentration. The error bars represent the standard deviations of the measured data for five samples.



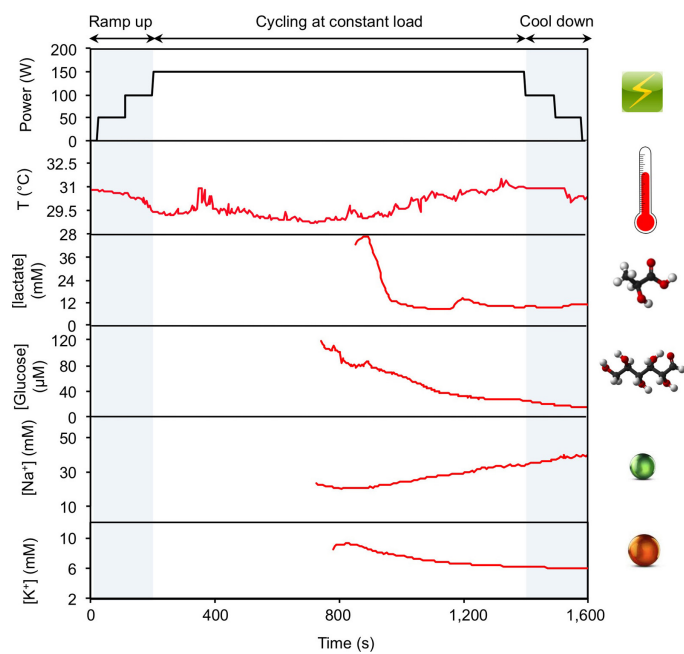
Extended Data Figure 7 | Selectivity study for electrochemical biosensors. **a–d**, The interference study for individual glucose (**a**), lactate (**b**), sodium (**c**) and potassium (**d**) sensors using an electrochemical working station. Data recording was paused for 30 s for the addition of each analyte in **c** and **d**. **e, f**, The real-time system-level interference study

(**e**) and calibration plot (**f**) of the amperometric glucose and lactate sensor array with a shared solid-state Ag/AgCl reference electrode. **g, h**, The real-time interference study (**g**) and calibration plot (**h**) of the potentiometric Na^+ and K^+ sensor array with a shared PVB-coated reference electrode. Data recording was paused for 30 s for each solution change in **e** and **g**.

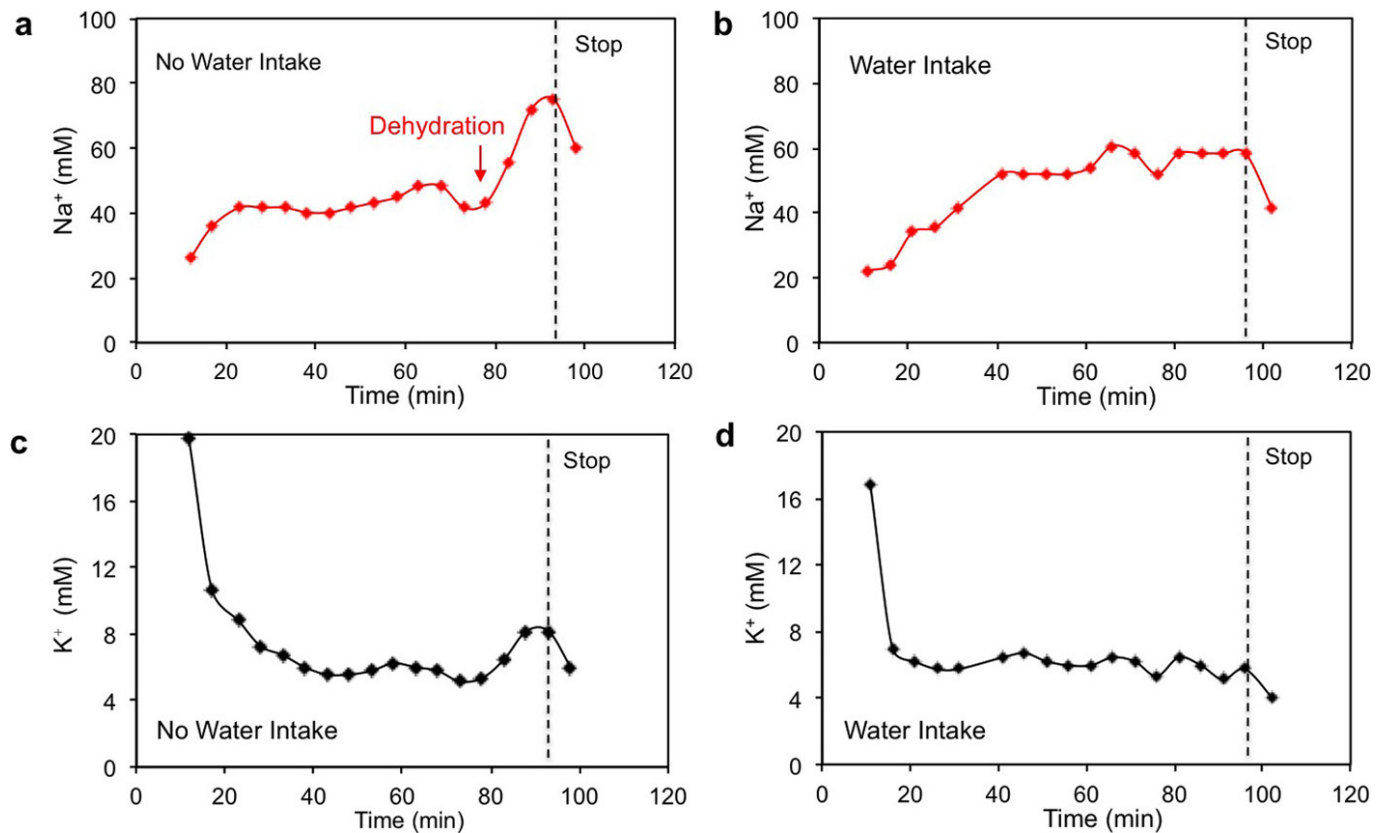


Extended Data Figure 8 | Mechanical deformation study of the flexible sensors and the FPCB. a–f, The responses of the sodium (a), potassium (b), glucose (c), lactate (d), temperature (e) sensors and of the FPCB (f) after 0, 30 and 60 cycles of bending. g–l, The responses of the sodium (g),

potassium (h), glucose (i), lactate (j), and temperature (k) sensors and of the FPCB (l) during bending. The radii of curvature for the bending study of sensors and the FPCB were 1.5 cm and 3 cm, respectively. Data recording was paused for 30 s to change the conditions and settings.



Extended Data Figure 9 | On-body real-time perspiration analysis during stationary cycling using the FISA on a subject's wrist. Conditions are as in Fig. 3c and d.



Extended Data Figure 10 | *Ex situ* measurement of collected sweat samples using the FISA on a subject during stationary cycling at 150 W. **a, c,** The *ex situ* results of $[\text{Na}^+]$ (**a**) and $[\text{K}^+]$ (**c**) from the sweat samples collected from the subject's forehead without water intake ($\sim 2.5\%$ of body

weight dehydration). **b, d,** The *ex situ* results of $[\text{Na}^+]$ (**b**) and $[\text{K}^+]$ (**d**) from the sweat samples collected from the subject's forehead with water intake (150 ml per 5 min).

Lithium-ion battery structure that self-heats at low temperatures

Chao-Yang Wang^{1,2}, Guangsheng Zhang¹, Shanhai Ge², Terrence Xu², Yan Ji², Xiao-Guang Yang¹ & Yongjun Leng¹

Lithium-ion batteries suffer severe power loss at temperatures below zero degrees Celsius, limiting their use in applications such as electric cars in cold climates and high-altitude drones^{1,2}. The practical consequences of such power loss are the need for larger, more expensive battery packs to perform engine cold cranking, slow charging in cold weather, restricted regenerative braking, and reduction of vehicle cruise range by as much as 40 per cent³. Previous attempts to improve the low-temperature performance of lithium-ion batteries⁴ have focused on developing additives to improve the low-temperature behaviour of electrolytes^{5,6}, and on externally heating and insulating the cells^{7–9}. Here we report a lithium-ion battery structure, the ‘all-climate battery’ cell, that heats itself up from below zero degrees Celsius without requiring external heating devices or electrolyte additives. The self-heating mechanism creates an electrochemical interface that is favourable for high discharge/charge power. We show that the internal warm-up of such a cell to zero degrees Celsius occurs within 20 seconds at minus 20 degrees Celsius and within 30 seconds at minus 30 degrees Celsius, consuming only 3.8 per cent and 5.5 per cent of cell capacity, respectively. The self-heated all-climate battery cell yields a discharge/regeneration power of 1,061/1,425 watts per kilogram at a 50 per cent state of charge and at minus 30 degrees Celsius, delivering 6.4–12.3 times the power of state-of-the-art lithium-ion cells. We expect the all-climate battery to enable engine stop–start technology capable of saving 5–10 per cent of the fuel for 80 million new vehicles manufactured every year¹⁰. Given that only a small fraction of the battery energy is used for self-heating, we envisage that the all-climate battery cell may also prove useful for plug-in electric vehicles, robotics and space exploration applications.

Figure 1a schematically shows a generic lithium (Li)-ion all-climate battery (ACB) cell. In addition to the three essential battery components—anode, cathode and electrolyte—we add here a fourth component: a nickel (Ni) foil 50 μm in thickness having two tabs, one at each end. Electrical resistance between the two tabs is designed to be 56 $\text{m}\Omega$ at room temperature (25 °C) to keep the cell voltage around 2 V and to avoid solid–electrolyte interphase decomposition and copper foil oxidation. One tab is electrically connected to the negative terminal, welded together with the tabs of all anode layers. The second tab of the Ni foil extends outside the cell to form a third terminal, the activation terminal, used to activate battery internal heating at low temperatures. A switch connects the activation terminal with the negative terminal. When the switch is left open during cell activation for self-heating, electrons must flow through the Ni foil, generating substantial ohmic heat, which rapidly warms up the core of the battery. Once the battery internal temperature reaches or exceeds 0 °C, thereby enabling the electrochemical interface to generate high power for both discharge and charge, the activation process is completed and the switch is closed. When the ACB cell operates at around room temperature, the switch between the activation terminal and negative terminal remains

closed, making electrons bypass the Ni foil and reverting the ACB cell to a conventional Li-ion cell with very low internal resistance and high power. The switch between activation terminal and negative terminal may be controlled by the cell surface temperature.

Figure 1b shows cell voltage and surface temperature evolutions during cell activation followed by a 1C discharge of a 7.5 amp-hour (Ah) ACB cell at –20 °C, and similar results are shown in Extended Data Fig. 1a and b for –30 °C and –40 °C, respectively. The cell exhibits sufficiently high voltage to stay ‘healthy’ (that is, the battery materials do not suffer potential degradation) throughout activation and 1C discharge processes even from as low as –40 °C. More noteworthy is that the cell surface temperature rises rapidly, in seconds, from an extremely cold environment to 0 °C within the activation process (better seen in the insets to Fig. 1b and Extended Data Fig. 1, where the cell activation process is magnified). It is clear that cell activation takes only 19.5 s, 29.6 s and 42.5 s from environments at –20 °C, –30 °C and –40 °C, respectively. After activation, the cell surface temperature drops slightly below the freezing point owing to large heat loss to the cold surroundings in these environmental-chamber tests; however, in reality it would remain around the freezing point owing to the thermal insulation usually applied around cells. The 1C discharge energy, calculated by integrating the area underneath each discharge curve, is 102 watt-hours per kilogram (Wh kg^{-1}) for the ACB cell at –40 °C, compared to only 0.3 Wh kg^{-1} for the baseline cell without Ni foil. The ACB cell thus provides much more usable energy, enabling a longer cruising range for an electric car, especially in extreme cold.

The ultrafast cell activation discovered in this work makes ACB cells technologically viable for boosting battery power. Fundamentally, the activation time may be estimated as follows. Assuming negligible heat loss from cell surfaces to the surroundings owing to a short time duration (this assumption is realistic as batteries are well insulated in vehicles), the energy balance during cell activation is:

$$\int_0^{\tau_{\text{act}}} I_{\text{act}}(U_0 - V_{\text{act}})dt = mc_p\Delta T \quad (1)$$

where I_{act} and V_{act} are current and output voltage during cell activation, U_0 is the thermal equilibrium potential of a Li-ion cell (~ 4.2 V for the cells used in this study), τ_{act} is activation time, m is cell mass, c_p is the specific heat of the cell, and ΔT is the rise in temperature from the initial ambient temperature to, for example, 0 °C. Assuming $c_p = 1,000 \text{ J kg}^{-1} \text{ K}^{-1}$ and using an average activation current I_{act} of 47.4 A in the –30 °C activation case (see Extended Data Fig. 2), the theoretical activation time τ_{act} is estimated to be 26.7 s, very close to the measured 29.6 s. This also indicates that the self-heating mechanism devised in the ACB cell structure is very energy-efficient ($\sim 90\%$ in this case). If $V_{\text{act}} = 0$ V activation is implemented, one can convert 10% more electric energy into internal heat for battery warm-up from very low temperatures, thereby further shortening activation time. This is the greatest advantage of ACB cells over existing battery heating methods, which are much more energy- and time-consuming^{7–9}. For example,

¹Department of Mechanical and Nuclear Engineering and Electrochemical Engine Center (ECEC), The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ²EC Power, 341 Science Park Road, State College, Pennsylvania 16803, USA.

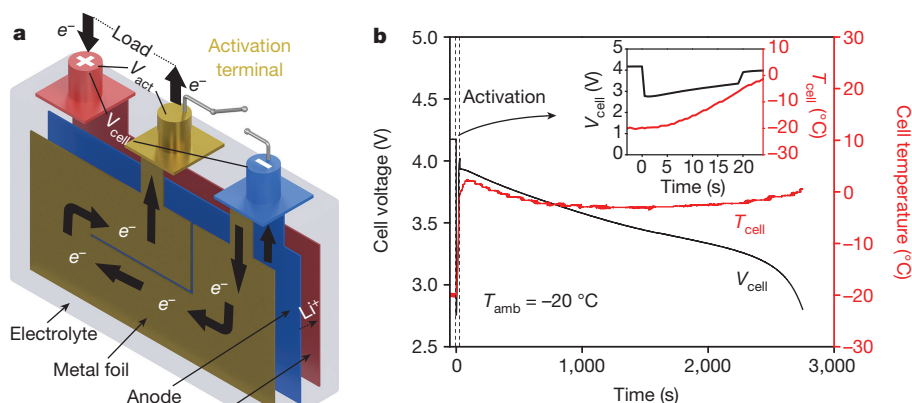


Figure 1 | The ACB. **a**, Schematic in which a metal foil is inserted to generate internal heating from a low temperature and to provide fast heat transfer to electrodes and electrolyte. This self-heating function is activated by turning off the switch between the activation terminal and the negative terminal. **b**, Cell voltage and temperature evolutions during

$V_{act} = 0.4\text{ V}$ activation (inset) and subsequent 1C discharge at -20°C . The battery temperature rises from -20°C to 0°C in $\sim 20\text{ s}$ and the 1C discharge thereafter occurs at the $\sim 0^\circ\text{C}$ battery core temperature rather than the -20°C ambient temperature.

Vlahinos and Pesaran⁷ computationally showed that battery core heating based on the cell's internal resistance is more effective than external heating methods. Stuart and Handeb⁸ argued that direct-current internal heating is ineffective and instead implemented expensive, heavy alternating-current generators for heating. More recently, Ji and Wang⁹ thoroughly reviewed a wide range of heating strategies for Li-ion batteries and demonstrated that self-resistive heating from -20°C to 20°C takes $\sim 120\text{ s}$ and consumes $\sim 15\%$ battery energy. For heating from -20°C to 0°C as in the present context, their cell would require a 60-s heating time and 7.5% energy consumption, much less efficient than the present ACB cell.

Another important feature of the ACB cell is high power, immediately available after ultrafast activation just as the battery materials and electrochemical interfaces reach 0°C . In Fig. 2a, for -20°C , -30°C and -40°C , a 10-s hybrid pulse power characterization (HPPC) power in watts per kilogram, for both discharge and regeneration (charge), as a function of depth of discharge is compared to that of a conventional Li-ion cell without Ni foil. At 50% state-of-charge (SOC) or depth of discharge, the power boost over the conventional Li-ion cell is 2.7, 6.4 and 25.1 for -20°C , -30°C and -40°C , respectively, for discharge, and 5.1, 12.3 and 55 for regeneration. Figure 2b plots the specific power versus ambient

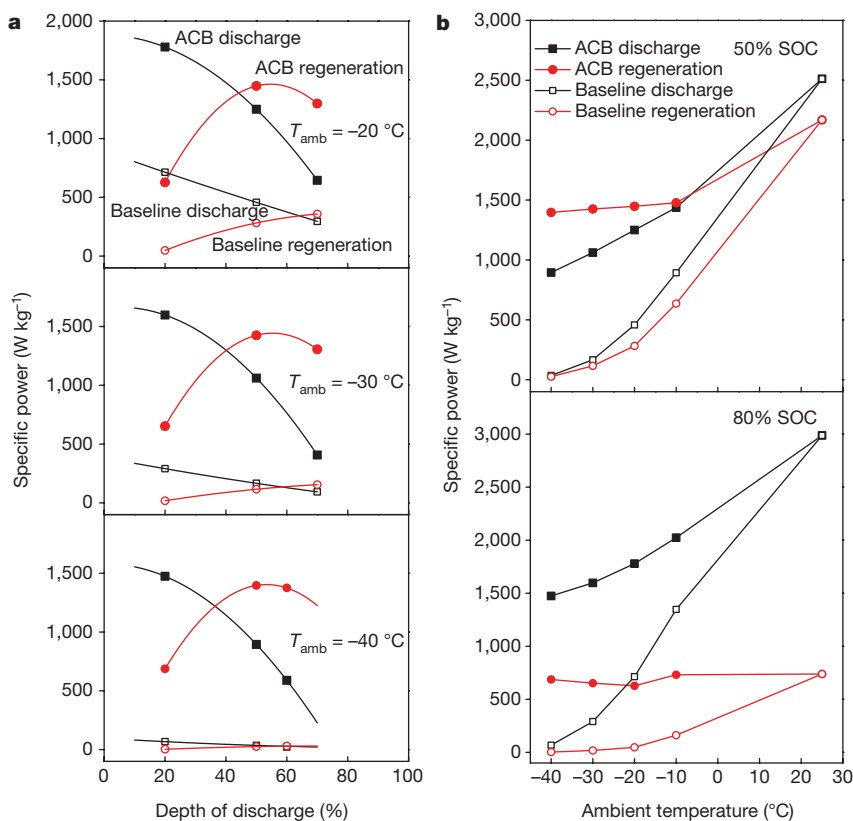


Figure 2 | Power performance of the ACB cell. **a**, 10-s HPPC specific power versus depth of discharge, compared to the baseline cell for -20°C , -30°C and -40°C . At 50% SOC, the ACB cell delivers 2.7 times, 6.4 times and 25.1 times the discharge power and 5.1 times, 12.3 times

and 55 times the regeneration power of a baseline cell at -20°C , -30°C and -40°C , respectively. **b**, 10-s HPPC specific power after activation versus the baseline as function of ambient temperature for 50% and 80% SOC.

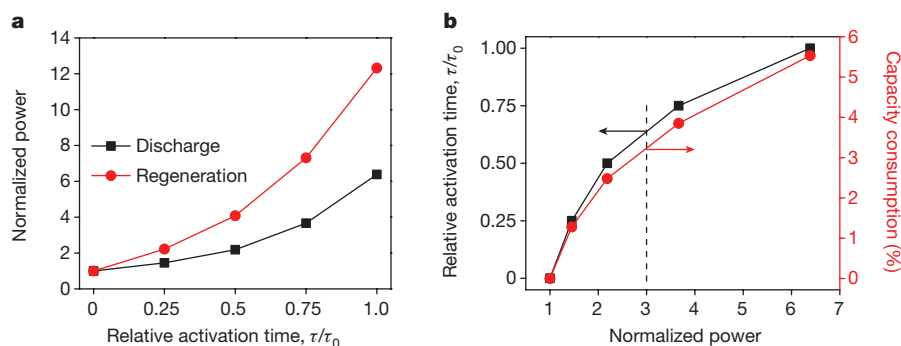


Figure 3 | Power on demand at 50% SOC for 10-s HPPC at -30°C . **a**, Normalized power (ACB/baseline) versus relative activation time. τ_0 is the time of full activation. **b**, Relative activation time and percentage

capacity consumption due to activation as functions of normalized power. 5.5% energy (the right y-axis) can be exchanged, on demand, for 640% power (the x-axis), or 3.2% energy can be exchanged for 300% power.

temperature for both ACB and baseline cells at 50% and 80% SOC, respectively. The discharge power (black lines in Fig. 2b) of the ACB cell is improved to $1,061\text{ W kg}^{-1}$ and $1,600\text{ W kg}^{-1}$ at -30°C for 50% and 80% SOC, respectively. These power levels are more than 5–6 times the power of the baseline Li-ion cell at the same temperature. Regeneration power at low temperatures is equally impressive for the ACB cell, reaching $1,425\text{ W kg}^{-1}$ at 50% SOC and 650 W kg^{-1} at 80% SOC at -30°C , indicative of unprecedented high charge/regeneration power in the extreme cold. These high power capabilities, readily available after a short activation, open new possibilities for a wide variety of applications where high battery power is critically sought. A few examples are the expedient capture of braking energy in extreme cold where it is most needed, weather-independent fast charging, and high-flying drones at low atmospheric temperatures.

For ACB cells to enjoy a dramatic power boost at low temperatures, some activation time and energy (or charge) consumption are required. Both can be further managed by exercising a power-on-demand strategy, that is, implementing partial activation to attain

a smaller but sufficient power boost. Such experiments are shown in Fig. 3a and b under the conditions of -30°C and 50% SOC, where normalized power, that is, the power of the ACB cell over the baseline, is plotted against the relative activation time. Obviously, for zero activation time or no activation at all, the normalized power is unity. At 100% activation, the normalized power for the ACB cell at -30°C reaches 6.4 times the discharge power and 12.3 times the regeneration power, respectively. However, at a partial activation such as 50%, there are already marked increases in both discharge and regeneration power, as can be seen from Fig. 3a. Figure 3b re-plots the relative activation time and capacity consumption percentage due to activation against the normalized power. It is seen that for a 3-fold power boost, 66% partial activation suffices and the capacity consumption due to cell activation is only 3.2% for heating from the ambient temperature of -30°C . Therefore, the power-on-demand strategy further reduces the activation time from 30 s to 20 s and also the capacity consumption due to activation from 5.5% to 3.2%, at the expense of having 3-fold power instead of 6.4-fold power.

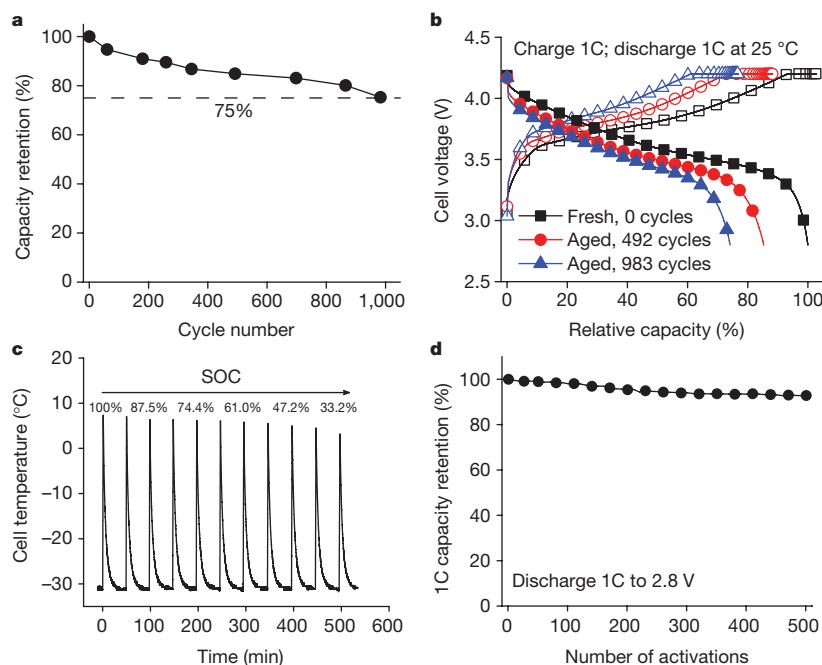


Figure 4 | ACB cell durability. **a**, C/3 capacity retention. In the present context, $C/3 = (7.5\text{ A})/3 = 2.5\text{ A}$ discharge current. **b**, 1C charge/discharge curves of fresh cells and cells aged from 45°C cycling between 2.8 V and 4.15 V. Both capacity retention and charge/discharge curves in **a** and **b** are obtained during cell characterization at 25°C . ACB cells give rise to almost no side effects in high-temperature cycling. **c**, Cell surface temperature versus time in a series of ten consecutive cycles of activation and cool-

down in durability of repetitive activations from $T_{\text{amb}} = -30^{\circ}\text{C}$. The change in SOC during the ten cycles is also indicated. **d**, 1C capacity versus number of activations for $T_{\text{amb}} = 25^{\circ}\text{C}$. The constant-current, constant-voltage charge protocol is constant current at 1C followed by constant voltage at 4.2 V and terminated when the charge current diminishes to C/20. Little degradation exists, even after 500 activations from -30°C .

To project a further large reduction in activation time, τ_{act} , and capacity consumption, Q_{act}/Q_0 , for future energy-dense electric-vehicle batteries, it follows from equation (1) that if the C-rate, β (which is the dimensionless electric current, relative to the cell capacity, such that in the present context, 1C for a 7.5-Ah cell is equivalent to a discharge current of 7.5 A), during activation is kept constant, the activation time and percentage capacity consumption are proportional to:

$$\tau_{\text{act}} = \frac{mc_p \Delta T}{Q_0(U_0 - V_{\text{act}})\beta} \propto \frac{\bar{V}}{U_0 - V_{\text{act}}} \frac{1}{\rho_e} \quad (2)$$

$$\frac{Q_{\text{act}}}{Q_0} = \frac{mc_p \Delta T}{Q_0(U_0 - V_{\text{act}})} \propto \frac{\bar{V}}{U_0 - V_{\text{act}}} \frac{1}{\rho_e} \quad (3)$$

where ρ_e is the cell's energy density and \bar{V} is the nominal voltage used in calculating the energy density.

Note that both activation time and capacity consumption will drop by half if the energy density is doubled from the current level of 170 Wh kg⁻¹ to a future level of 340 Wh kg⁻¹. The halved activation time and capacity consumption would be at levels of ~ 10 s and 1.9% for ACB cell activation from -20°C , a common low-temperature environment. Further implementing a partial activation based on power demand, it would be possible to keep the ACB activation time within 5 s and capacity consumption within 1%, while still delivering sufficient power for a wide range of applications involving cold climates.

Existing techniques improve low-temperature power at the expense of greatly deteriorated performance and lifespan at high temperatures. Figure 4a and b shows no additional side effects with high-temperature cycling, and a normal 17% loss of capacity with room-temperature cycling over 2,000 cycles (Extended Data Fig. 3).

Finally, we explore cell degradation caused by repetitively activating an ACB cell from -30°C followed by forced air cool-down. A consecutive ten cycles of activation and cool-down is carried out for an ACB cell, starting with 100% SOC (Fig. 4c). Thereafter, the cell is charged back to 100% SOC at room temperature and with standard constant-current, constant-voltage protocol having a voltage limit of 4.2 V. A total of 500 such activations are carried out, and the cell capacity at room temperature and 1C rate characterized at the end of every ten activation/cool-down cycles is shown in Fig. 4d. The cell capacity fade is less than 7.2% at the end of 500 activations. In a practical electric-vehicle battery pack, once all batteries are heated to a higher temperature, the cool-down timescale usually ranges from several hours to 10–15 h. This implies that cell activation is probably only needed once per day. Assuming 30 days of extreme weather (-30°C) per year, 500 activations tested in the durability experiment shown in Fig. 4d would be equivalent to about 16 years of operation, meaning battery life is not noticeably decreased by cell activation from subfreezing temperatures.

The new material we added into the baseline battery to make it an ACB cell—that is, the Ni foil—weighs about 100 g per kilowatt-hour battery and costs US\$0.1 per kilowatt-hour based upon a Ni price of US\$10 per kilogram. Compared to the current best specific energy of Li-ion battery systems, which is 150 Wh per kilogram of the battery system, and assuming a battery cost of US\$250 per kilowatt-hour (ref. 11), the added weight and cost due to ACB technology are 1.5% and 0.04% of those of the baseline battery.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 September; accepted 30 November 2015.

Published online 20 January 2016.

- Armand, M. & Tarascon, J. M. Building better batteries. *Nature* **451**, 652–657 (2008).
- Villasenor, J. High-altitude surveillance drones: coming to a sky near you. *Sci. Am.* **Feb**, 24 (2012).
- Extreme temperatures affect electric vehicle driving range, AAA says. <http://newsroom.aaa.com/2014/03/extreme-temperatures-affect-electric-vehicle-driving-range-aaa-says/> *Newsroom* (20 March 2014).
- Ji, Y., Zhang, Y. & Wang, C. Y. Li-ion cell operation at low temperatures. *J. Electrochem. Soc.* **160**, A636–A649 (2013).
- Zhang, S. S., Xu, K. & Jow, T. R. A new approach toward improved low temperature performance of Li-ion battery. *Electrochem. Commun.* **4**, 928–932 (2002).
- Smart, M. C., Whitacre, J. F., Ratnakumar, B. V. & Amine, K. Electrochemical performance and kinetics of $\text{Li}_{1-x}(\text{Co}_{1/3}\text{Ni}_{1/3}\text{Mn}_{1/3})_{1-x}\text{O}_2$ cathodes and graphite anodes in low-temperature electrolytes. *J. Power Sources* **168**, 501–508 (2007).
- Vlahinos, A. & Pesaran, A. A. Energy efficient battery heating in cold climates. Society of Automotive Engineers (SAE) Technical Paper 2002-01-1975, <http://papers.sae.org/2002-01-1975/> (SAE, 2002).
- Stuart, T. A. & Handeb, A. HEV battery heating using AC currents. *J. Power Sources* **129**, 368–378 (2004).
- Ji, Y. & Wang, C. Y. Heating strategies for Li-ion batteries operated from subzero temperatures. *Electrochim. Acta* **107**, 664–674 (2013).
- Chen, K. et al. Evaluation of the low temperature performance of lithium manganese oxide/lithium titanate lithium-ion batteries for start/stop applications. *J. Power Sources* **278**, 411–419 (2015).
- Gröger, O., Gasteiger, H. A. & Suchsland, J.-P. Electromobility: batteries or fuel cells? *J. Electrochem. Soc.* **162**, A2605–A2622 (2015).

Acknowledgements We thank W. Zhao and C. E. Shaffer for early discussions on using battery simulation software to discover the all-climate battery. This work was inspired by US patent publication numbers 2014-0342194, 2015-0303444 and 2015-0104681 and Patent Cooperation Treaty publication numbers WO 2014/186195, WO 2015/102709 and WO 2015/102708.

Author Contributions C.Y.W. developed the concept and wrote the manuscript. S.G., T.X., Y. J. and X.G.Y. designed and built the cells, G.Z. built the test stand and carried out the performance characterization, and Y.L. performed the cycle life experiments. All authors contributed to development of the manuscript and to discussions as the project developed.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.Y.W. (cxw31@psu.edu or cywang@ecpouergroup.com).

METHODS

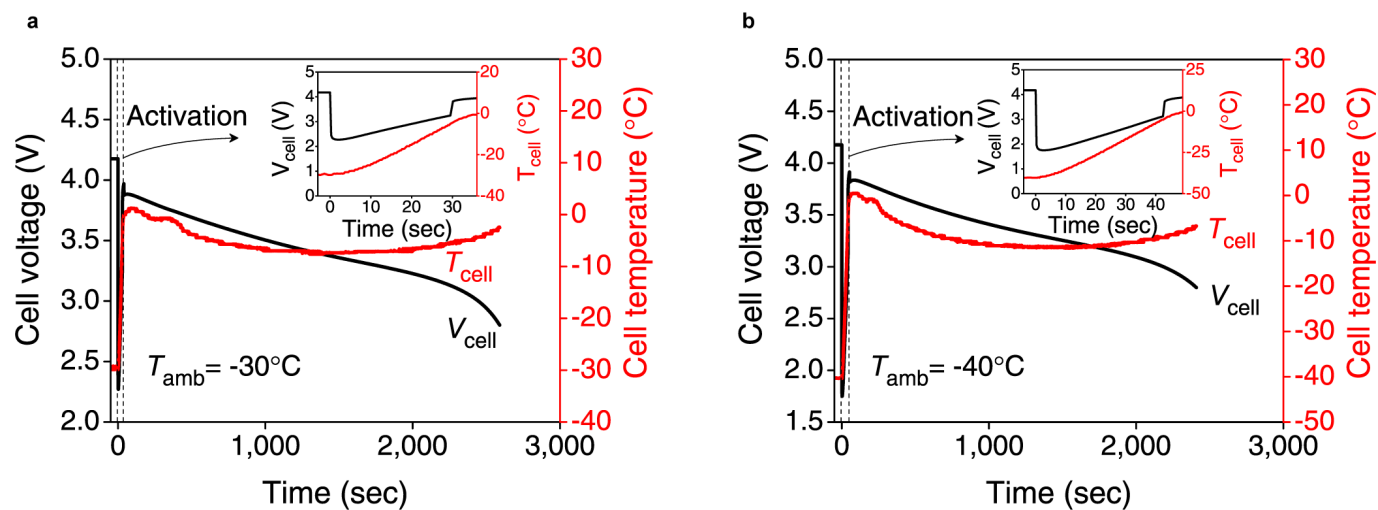
We fabricate 7.5-Ah ACB pouch cells using $\text{LiNi}_{0.5}\text{Co}_{0.2}\text{Mn}_{0.3}\text{O}_2$ (Umicore) as cathodes and graphite (Nippon Carbon) as anodes with 1 M of LiPF_6 dissolved in ethylene carbonate/ethyl methyl carbonate (3:7 by weight) + 2% vinylene carbonate as electrolyte (materials from BASF). The capacity ratio of negative to positive electrode is designed to be 1.2. The 7.5-Ah pouch cell contains a stack of 26 anode and 25 cathode layers. A Celgard-2325 separator of thickness $25\text{ }\mu\text{m}$ is used. A Ni foil sized at $56\text{ m}\Omega$ at room temperature is coated with a thin backing material of polyethylene terephthalate ($28\text{ }\mu\text{m}$) for electrical insulation and sandwiched between two single-sided anode layers and the three-layer assembly then stacked in the centre of the cell.

The cathodes are prepared by coating *N*-methylpyrrolidone-based slurry onto $15\text{-}\mu\text{m}$ -thick Al foil, whose dry material consists of NCM523 (92 wt%), Super-P (Timcal) (4 wt%) and polyvinylidene fluoride (Arkema) (4 wt%) as a binder. The anodes are prepared by coating deionized water-based slurry onto $10\text{-}\mu\text{m}$ -thick Cu foil, whose dry material consists of graphite (97.5 wt%), styrene butadiene rubber (Zeon) (1.5 wt%) and carboxymethyl cellulose (Dai-Ichi Kogyo Seiyaku) (1 wt%).

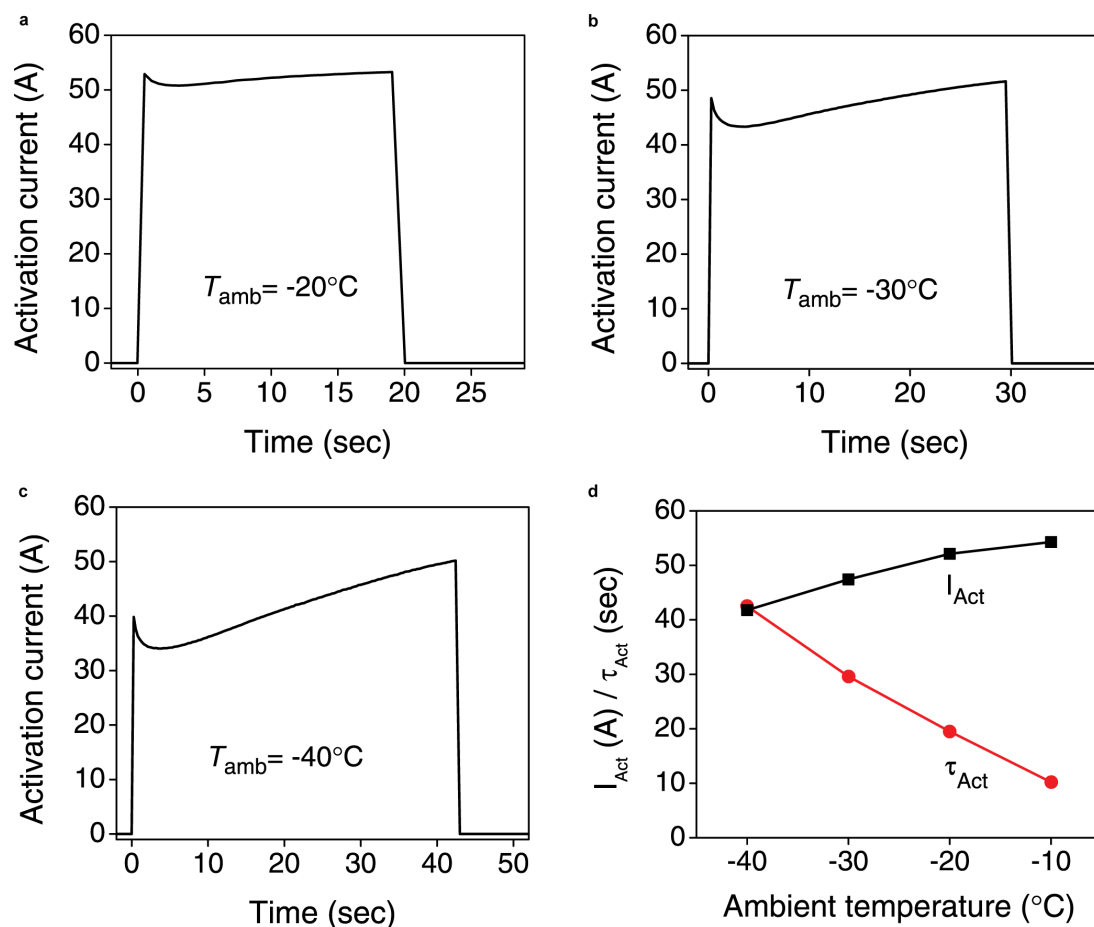
Each ACB pouch cell has a $152\text{ mm} \times 75\text{ mm}$ footprint area, weighs 160 g, and has a nominal capacity of 7.5 Ah with a specific energy of 170 Wh kg^{-1} and an energy density of 327 Wh per litre . The discharge performance of the ACB cell at room temperature without activation is shown in Extended Data Fig. 4 as a function of the C-rate.

We denote the voltage between the positive and negative terminals as cell voltage, a potential window encompassing all battery materials. Additionally, we denote the voltage between the positive and activation terminals as V_{act} for the activation process only. For any subzero operation, cell activation is first carried out by a constant-voltage, constant-current protocol where constant voltage means that V_{act} is set at 0.4 V until the current reaches and is limited at 60 A (that is, 8C). Cell activation is terminated when the cell temperature reaches $-5\text{ }^\circ\text{C}$ as measured by a thermocouple placed at the centre of the cell's outer surface. A 10-s rest is given between the end of activation and cell loading for equilibrium, during which the cell surface temperature usually continues to rise to $0\text{ }^\circ\text{C}$. Hence, cell activation described in the present work is designed to bring the battery core temperature to or above the freezing point from any subzero ambient environment. Prior to any subfreezing tests, an ACB cell is soaked in the environmental chamber for 8–12 h to reach thermal equilibrium with the ambient temperature.

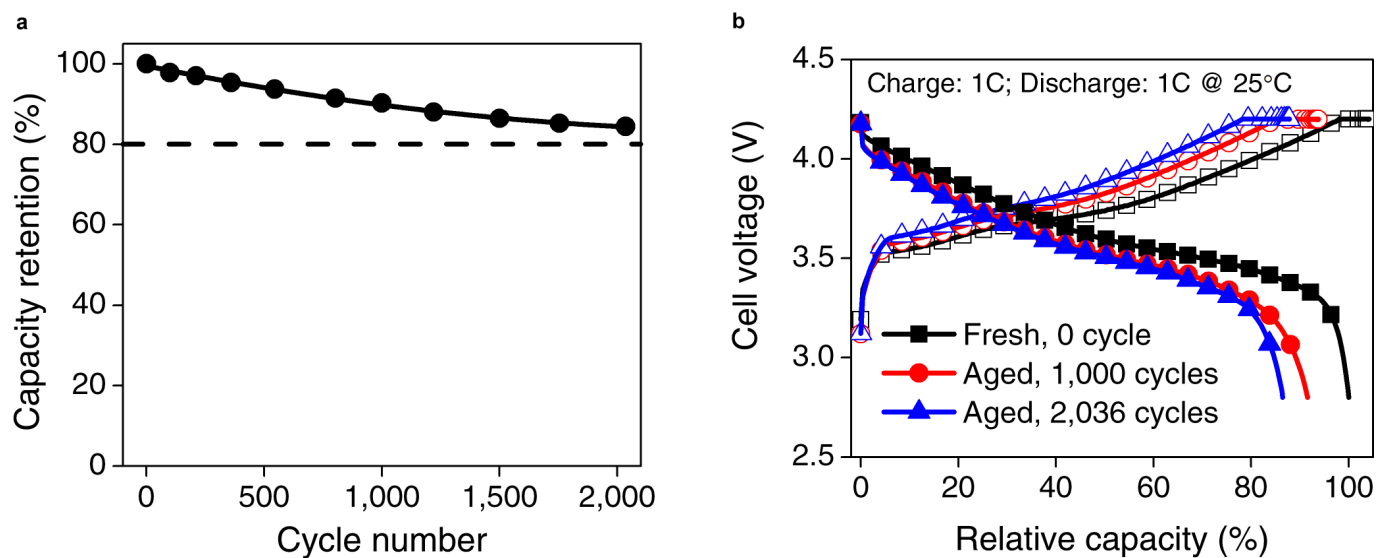
Two types of cell discharge are performed in the present work. One is 1C discharge with a cutoff voltage of 2.8 V and the other is 10-s HPPC in which at a given SOC level, a 10-s charge pulse is applied at $V_{\text{max}} = 4.2\text{ V}$, followed by a 40-s rest and a discharge pulse at $V_{\text{min}} = 2.8\text{ V}$. The discharge and charge (or regeneration) power, in watts per kilogram of the battery cell is calculated as the product of constant voltage and average current in the 10-s discharge and charge pulses, then divided by the cell weight.



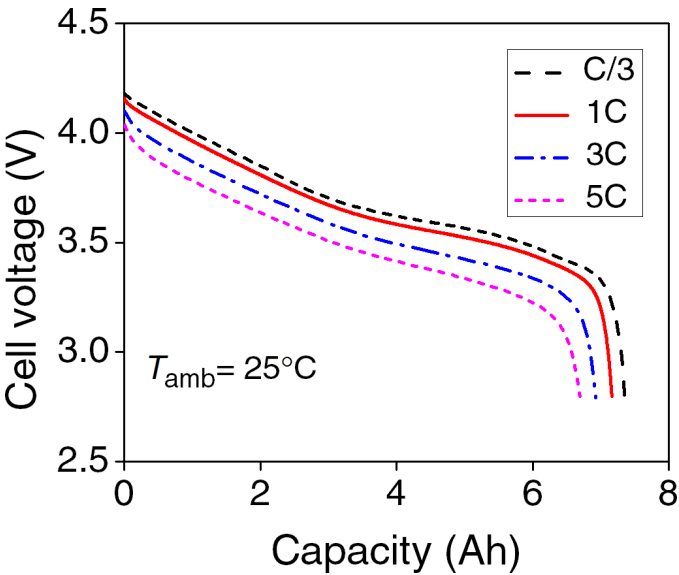
Extended Data Figure 1 | Cell voltage and temperature evolution during activation and subsequent 1C discharge. a, -30°C . b, -40°C . The insets show the $V_{\text{act}} = 0.4\text{ V}$ activation more clearly.



Extended Data Figure 2 | Cell current variations during activation. a, -20°C . b, -30°C . c, -40°C . d, Activation time τ_{act} and average activation current I_{act} versus the ambient temperature T_{amb} .



Extended Data Figure 3 | 1C charge/2C discharge cycling of ACB cell at room temperature between 2.8 V and 4.2 V. a, C/3 capacity retention. b, 1C charge/discharge curves of the fresh and aged cells.



Extended Data Figure 4 | ACB cell discharge with various C-rates of discharge and at room temperature.

No iron fertilization in the equatorial Pacific Ocean during the last ice age

K. M. Costa^{1,2}, J. F. McManus^{1,2}, R. F. Anderson^{1,2}, H. Ren³, D. M. Sigman⁴, G. Winckler^{1,2}, M. Q. Fleisher¹, F. Marcantonio⁵ & A. C. Ravelo⁶

The equatorial Pacific Ocean is one of the major high-nutrient, low-chlorophyll regions in the global ocean. In such regions, the consumption of the available macro-nutrients such as nitrate and phosphate is thought to be limited in part by the low abundance of the critical micro-nutrient iron¹. Greater atmospheric dust deposition² could have fertilized the equatorial Pacific with iron during the last ice age—the Last Glacial Period (LGP)—but the effect of increased ice-age dust fluxes on primary productivity in the equatorial Pacific remains uncertain^{3–6}. Here we present meridional transects of dust (derived from the ²³²Th proxy), phytoplankton productivity (using opal, ²³¹Pa/²³⁰Th and excess Ba), and the degree of nitrate consumption (using foraminifera-bound $\delta^{15}\text{N}$) from six cores in the central equatorial Pacific for the Holocene (0–10,000 years ago) and the LGP (17,000–27,000 years ago). We find that, although dust deposition in the central equatorial Pacific was two to three times greater in the LGP than in the Holocene, productivity was the same or lower, and the degree of nitrate consumption was the same. These biogeochemical findings suggest that the relatively greater ice-age dust fluxes were not large enough to provide substantial iron fertilization to the central equatorial Pacific. This may have been because the absolute rate of dust deposition in the LGP (although greater than the Holocene rate) was very low. The lower productivity coupled with unchanged nitrate consumption suggests that the subsurface major nutrient concentrations were lower in the central equatorial Pacific during the LGP. As these nutrients are today dominantly sourced from the Subantarctic Zone of the Southern Ocean, we propose that the central equatorial Pacific data are consistent with more nutrient consumption in the Subantarctic Zone, possibly owing to iron fertilization as a result of higher absolute dust fluxes in this region^{7,8}. Thus, ice-age iron fertilization in the Subantarctic Zone would have ultimately worked to lower, not raise, equatorial Pacific productivity.

The major nutrients for phytoplankton growth (nitrogen, phosphorus and silicon) are supplied to the surface waters of the equatorial Pacific by wind-driven upwelling along the Equator. Their consumption by phytoplankton is thought to be limited in part by the low concentrations of the critical micro-nutrient iron¹. Successful iron fertilization experiments in the modern ocean⁹ have demonstrated the sensitivity of these regions to changes in the micro-nutrient supply. Dust dissolution is one source of iron to the ocean, and globally increased dust fluxes² may have caused natural iron fertilization during the peak of the LGP. There is evidence for iron fertilization⁸ in the Subantarctic Zone of the Southern Ocean, and the associated carbon storage in the deep ocean may have been responsible for almost half of the carbon dioxide drawdown during the LGP¹⁰. However, the effects of increased ice-age dust fluxes on the equatorial Pacific are debated^{3–6}, with arguments both for and against iron fertilization, particularly in the eastern equatorial Pacific.

Here we present new proxy data on dust flux (²³²Th flux, see Methods), biological productivity ('export production', the export of organic matter out of surface water, as reconstructed from the opal flux, excess barium flux, and ²³¹Pa/²³⁰Th, for which ²³⁰Th and ²³¹Pa represent excess initial ²³⁰Th and ²³¹Pa, respectively) and the degree of nitrate consumption (foraminifera-bound $\delta^{15}\text{N}$) from a north–south transect of six cores from the central equatorial Pacific (0.22°S to 6.83°N, 156°–161°W; Extended Data Fig. 1) at two time slices: the Holocene (0–10,000 years ago) and the LGP (17,000–27,000 years ago). The relatively shallow water depths (average ~3,000 m) result in low rates of carbonate dissolution and permit the development of robust foraminifera-based radiocarbon age models (Extended Data Fig. 2, Extended Data Table 1). Furthermore, these core sites are far from the eastern continental margins, and so ²³²Th at these sites predominantly reflects the flux of airborne dust particles². Central equatorial Pacific surface waters are dominantly sourced with nitrate from the Equatorial Undercurrent, which originates in the west¹¹. Thus, relative to the tropical Pacific as whole, the $\delta^{15}\text{N}$ of the nitrate supply in the central equatorial Pacific is unlikely to be particularly sensitive to changes in eastern Pacific denitrification (see Methods).

Because the central equatorial Pacific is far from dust sources, reconstructed dust fluxes are among the lowest ever measured¹². Core-top dust fluxes along the 160°W transect average 11.0 mg cm⁻² kyr⁻¹, with a maximum of 12.8 mg cm⁻² kyr⁻¹ at 2.46°N (Fig. 1a). There is a weak decline in dust flux with increasing latitude ($r^2 = 0.42$, $P = 0.88$), with the lowest dust flux (8.8 mg cm⁻² kyr⁻¹) at the most northerly core. This negative correlation is in contrast to more easterly (110°W, 140°W) meridional transects, where the highest dust fluxes occur at the more northerly cores¹³. Relative to the Holocene, ice-age dust fluxes are two to three times greater along the 160°W transect, averaging 28.6 mg cm⁻² kyr⁻¹, with a maximum of 32.2 mg cm⁻² kyr⁻¹, at 2.46°N. The dust fluxes are remarkably constant as a function of latitude. Overall, the greater dust fluxes during the LGP are consistent with other reconstructions across the equatorial Pacific, which find glacial dust fluxes 0.7 to 3.4 times those of the Holocene (Fig. 2).

However, the expectations of ice-age iron fertilization do not correspond with the observed changes in surface productivity (as determined from opal flux, excess barium flux, ²³¹Pa/²³⁰Th; see Methods). Core-top opal fluxes along the transect at 160°W average 47 mg cm⁻² kyr⁻¹ and are negatively correlated with latitude ($r^2 = 0.90$, $P = 0.31$) (Fig. 1b). The maximum opal flux (70 mg cm⁻² kyr⁻¹) occurs at the Equator, which is consistent with higher surface productivity within the equatorial upwelling zone. Compared to the core-top fluxes, glacial opal fluxes are mostly lower, averaging 37 mg cm⁻² kyr⁻¹, a finding that is inconsistent with the expectations of local iron fertilization. Glacial fluxes also diminish northward from the Equator, consistent with a stable position for the upwelling.

¹Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA. ²Department of Earth and Environmental Sciences, Columbia University, New York, New York 10027, USA. ³Department of Geosciences, National Taiwan University, Taipei 106, Taiwan. ⁴Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA. ⁵Department of Geology and Geophysics, Texas A&M University, College Station, Texas 77843, USA. ⁶Ocean Sciences Department, University of California, Santa Cruz, California 95064, USA.

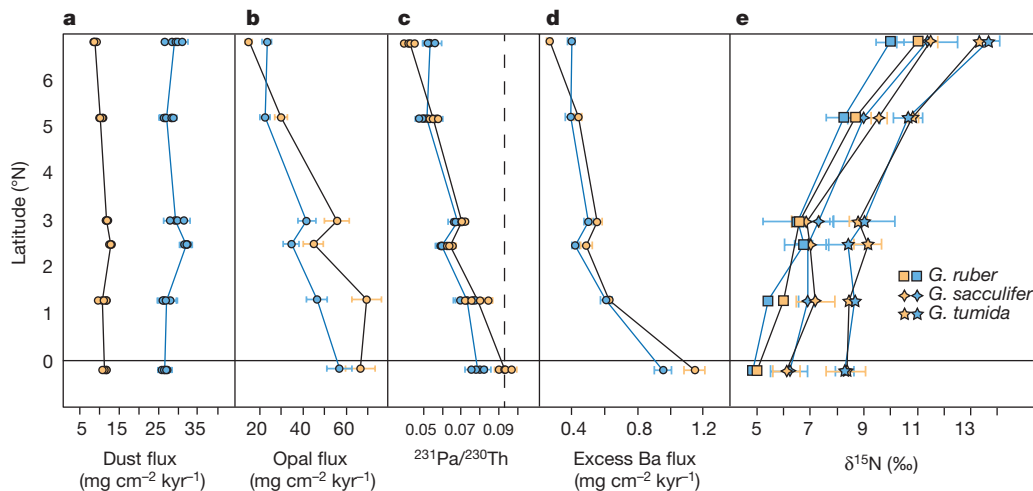


Figure 1 | ^{230}Th -normalized dust flux (see Methods), opal flux, $^{231}\text{Pa}/^{230}\text{Th}$, excess Ba flux, and FB- $\delta^{15}\text{N}$. **a–e**, Samples from the Holocene (0–10,000 years ago) are orange; samples from the LGP (17,000–27,000 years ago) are blue. The magnitude of variability in dust fluxes within the Holocene ($n = 30$) and within the LGP ($n = 31$)

is negligible compared to the more than doubling of dust fluxes in the LGP compared to the Holocene. Error bars are 2σ and indicate analytical precision. Error bars on FB- $\delta^{15}\text{N}$ reflect the variation of samples within the time slice (indicating 2σ). Data for this figure are given in the Source Data for Fig. 1.

The glacial–interglacial productivity signal is corroborated by initial $^{231}\text{Pa}/^{230}\text{Th}$, an opal flux proxy that has the advantage of being insensitive to remineralization¹⁴, and excess barium (Ba) fluxes, an independent proxy for total export production¹⁵. Relative to the Holocene, $^{231}\text{Pa}/^{230}\text{Th}$ ratios and excess Ba fluxes are generally lower during the LGP (Fig. 1c, d), except for the northernmost site outside the upwelling zone. The consistent latitudinal trends, with the highest productivity at the Equator and decreasing productivity northward, suggest no major difference in the upwelling regime between the Holocene and the LGP. Furthermore, measured $^{231}\text{Pa}/^{230}\text{Th}$ ratios are constant throughout each time slice (Extended Data Fig. 3), indicating that our results are not biased by bioturbation of transient extreme values

(for example, a deglacial productivity maximum¹⁶). The positive correlation between the excess Ba fluxes, $^{231}\text{Pa}/^{230}\text{Th}$, and opal fluxes (Extended Data Fig. 6) demonstrates that, in this region, siliceous productivity is a proxy for total export production. Accordingly, both core-top and glacial $^{231}\text{Pa}/^{230}\text{Th}$ from across the equatorial Pacific^{3,17} (Fig. 2) indicate that lower surface productivity was a regional phenomenon during the LGP.

Across much of the global ocean, including most of the tropics and subtropics, the supply of major nutrients limits productivity. Iron limitation generally applies in regions with upwelling or deep mixing, where the major nutrient supply is increased but (given the low iron-to-major-nutrient concentration ratios of most deep waters) iron

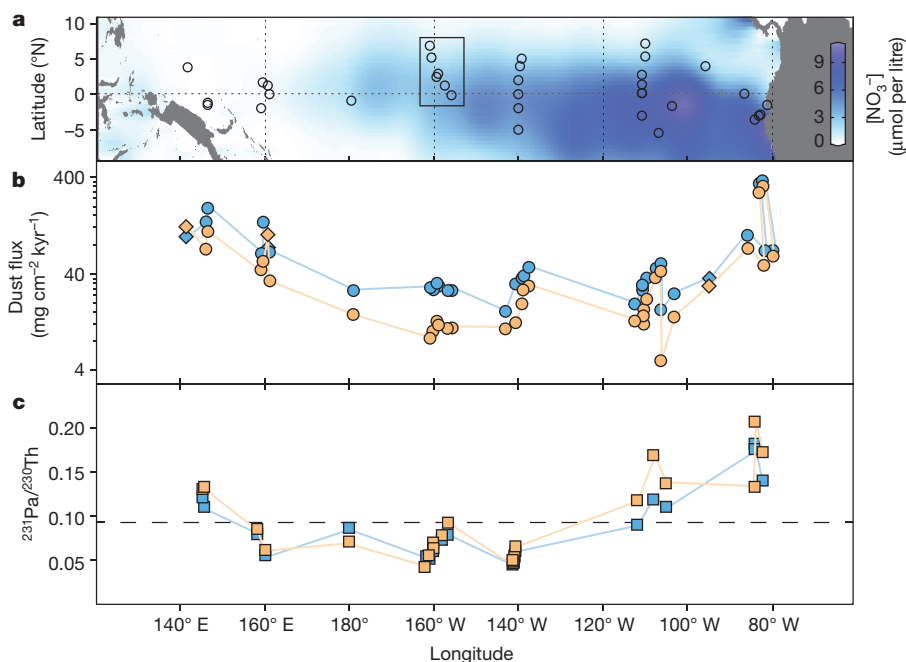


Figure 2 | Dust flux and $^{231}\text{Pa}/^{230}\text{Th}$ across the equatorial Pacific. **a**, Map of annual average surface nitrate concentrations³⁰. Black circles indicate the core locations, and the black box identifies the six new cores presented in this study. **b**, Dust flux inferred from ^{230}Th -normalized ^{232}Th fluxes (circles) and lithogenic accumulation rates (diamonds) for the Holocene (0–10,000 years ago, orange symbols) and the LGP (17,000–27,000 years

ago, blue symbols). Meridional transects at 110°W and 140°W are offset longitudinally to clarify the Holocene and LGP time slices. **c**, $^{231}\text{Pa}/^{230}\text{Th}$ for the Holocene (orange squares) and the LGP (blue squares). The dashed line represents the production ratio (0.093; see Methods). Data and references for this figure are given in the Source Data for Fig. 2.

supply is not equivalently increased, leading to incomplete consumption of the major nutrients. Therefore, iron fertilization by dust could drive an increase in the degree of major nutrient consumption in the upwelled surface waters of the equatorial Pacific. The nitrogen isotopic ratio of organic matter bound within the shell walls of planktonic foraminifera (foraminifera-bound $\delta^{15}\text{N}$, FB- $\delta^{15}\text{N}$) covaries with the degree of nitrate consumption⁸, and here we present FB- $\delta^{15}\text{N}$ data from two euphotic-zone species (*Globigerinoides ruber* and *G. sacculifer*) and one deeper-dwelling species (*G. tumida*). In both the Holocene and the LGP, FB- $\delta^{15}\text{N}$ differences among the species parallel depth habitat (Fig. 1e), consistent with previous findings of higher FB- $\delta^{15}\text{N}$ in deeper-dwelling species¹⁸. For each species, FB- $\delta^{15}\text{N}$ increases northward from the Equator, reflecting the progressive consumption of nitrate as surface waters flow away from the centre of upwelling, with the FB- $\delta^{15}\text{N}$ rise accelerating north of 3° N, as expected when nitrate consumption approaches completion¹⁹. For each species, core-top and glacial FB- $\delta^{15}\text{N}$ are remarkably similar across the meridional transect of cores. The lack of any difference between LGP and Holocene values argues strongly against glacial iron fertilization, which would be expected both to raise FB- $\delta^{15}\text{N}$ over the equatorial upwelling and to steepen the northward FB- $\delta^{15}\text{N}$ increase.

The unchanged relationship between $^{231}\text{Pa}/^{230}\text{Th}$ and opal flux (Extended Data Fig. 6) also supports sustained iron limitation. Silicification of diatom frustules is thought to be sensitive to iron stress⁶, with more robust and, therefore, better preserved frustules produced under conditions of greater iron limitation. The constancy of the relationship between opal flux and $^{231}\text{Pa}/^{230}\text{Th}$ (Extended Data Fig. 6), which would be sensitive to changes in opal preservation, indicates that any change in silicification due to iron stress between the LGP and Holocene was too small to alter this relationship measurably.

The lack of ice-age iron fertilization in the equatorial Pacific contrasts with its evident occurrence in the Southern Ocean^{7,8}. This dissimilarity may be due to three different factors related to dust fluxes and their fertilizing effects. First, the absolute fluxes of dust to the equatorial Pacific are so low in the Holocene that a doubling or even tripling of these fluxes during the LGP would still have amounted to only a minor atmospheric input of iron. Ice-age dust fluxes in the Subantarctic Zone of the Southern Ocean ($500\text{--}700\text{ mg cm}^{-2}\text{ kyr}^{-1}$)⁷ were 20–30 times greater than ice-age dust fluxes in the central equatorial Pacific ($20\text{--}30\text{ mg cm}^{-2}\text{ kyr}^{-1}$), suggesting that glacial dust in the equatorial Pacific remained roughly an order of magnitude too low to relieve iron limitation. Second, iron in equatorial Pacific dust may be less biologically accessible than the iron in Southern Ocean dust, given that recent work on mineralogical solubility of dust indicates that dust derived from glacial till (as was deposited in the ice-age Southern Ocean) is two to three times more soluble than dust derived from arid soils (as was deposited in the equatorial Pacific)²⁰. Third, the concentration of iron upwelled from the Equatorial Undercurrent, the principal regional iron source²¹, is augmented by partial dissolution of fluvial particles eroded from Papua New Guinea²². Regionally dry conditions during the LGP²³ may have reduced erosion and river transport rates, resulting in lower iron concentrations in the Equatorial Undercurrent that may have countered the small increase in atmospheric deposition of iron, although a full investigation into this negative feedback is beyond the scope of this study.

The combination of lower productivity and the constant degree of nitrate consumption suggests that the nitrate supply in the equatorial Pacific was lower in the LGP than in the Holocene. Reduced nitrate supply could be the result of (1) lower nitrate concentrations in upwelling waters, (2) a deeper thermocline, and (3) lower upwelling rates. Coupled sea surface temperature and productivity records from the eastern equatorial Pacific argue against large differences in upwelling rates between the LGP and the late Holocene²⁴. Moreover, in the modern central equatorial Pacific, the degree of nitrate consumption is observed to rise during the springtime minimum in upwelling

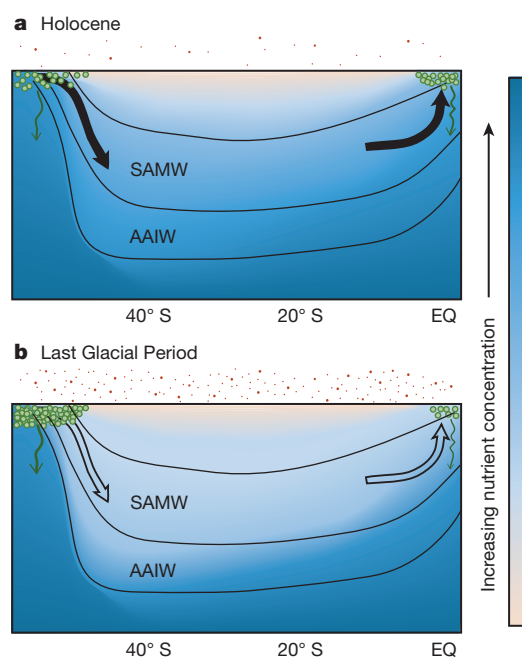


Figure 3 | Changes in nutrient dynamics between the Holocene and the LGP. General circulation shown as black arrows. AAIW, Antarctic Intermediate Water. During the LGP, a 4–6-fold increase in dust supply (red circles) stimulated biological productivity (green circles and arrows) and enhanced the degree of nutrient consumption in the SAMW formation regions to levels much greater than during the Holocene^{7,8}. Greater glacial productivity left a smaller inventory of nutrients to be subducted into the thermocline during SAMW formation, thereby lowering the supply of nutrients to equatorial upwelling regions.

rate¹¹, and palaeoceanographic data suggest that it also rose during orbitally driven phases of lower upwelling²⁵.

Conversely, there is evidence to support both lower major nutrient concentrations in the water upwelled in the equatorial Pacific²⁶ and a deeper thermocline²⁷ during the LGP. The aforementioned dust-driven enhancement of nitrate consumption in Subantarctic Zone surface waters would have reduced the nitrate concentrations exported in Subantarctic Mode Water (SAMW) to the low latitudes²⁸ (Fig. 3). These lower nitrate concentrations in upwelled waters can explain a decline in productivity in spite of a constant degree of nitrate consumption, especially if combined with a deeper thermocline, which would inhibit the entrainment of nutrient-rich waters into the upwelling zone. Furthermore, the constant degree of nitrate consumption suggests that the iron content of the upwelled water was lower during the LGP by the same proportion that major nutrients were lower relative to Holocene concentrations. This parallel response may indicate that the iron content of shallow subsurface waters maintains a relatively constant ratio with the major nutrients, through a conserved iron-to-major-nutrient ratio in sinking organic matter²⁹.

This study demonstrates that there was no enhancement of nitrate uptake by dust (iron) fertilization in the equatorial Pacific during the LGP. However, glacial–interglacial changes in dust flux do affect the equatorial Pacific through changes in global nutrient dynamics. Iron fertilization in the Southern Ocean during the LGP generated a higher degree of nutrient consumption, higher surface productivity, and lower nutrient export in the SAMW-formation regions²⁸, which reduced the nutrient supply to the equatorial Pacific compared to the Holocene. The inverse correlation between the Subantarctic Zone of the Southern Ocean and equatorial Pacific surface productivity on glacial–interglacial timescales implies that increases in Southern Ocean productivity occur at the expense of equatorial Pacific productivity, a compensating mechanism by which globally integrated ocean productivity would be stabilized over glacial cycles.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 February; accepted 9 November 2015.

- Brzezinski, M. A. *et al.* Co-limitation of diatoms by iron and silicic acid in the equatorial Pacific. *Deep Sea Res. Part II* **58**, 493–511 (2011).
- Winckler, G., Anderson, R. F., Fleisher, M. Q., McGee, D. & Mahowald, N. Covariant glacial-interglacial dust fluxes in the equatorial Pacific and Antarctica. *Science* **320**, 93–96 (2008).
- Bradtmeier, L. I., Anderson, R. F., Fleisher, M. Q. & Burckle, L. H. Diatom productivity in the equatorial Pacific Ocean from the last glacial period to the present: a test of the silicic acid leakage hypothesis. *Paleoceanography* **21**, PA4201 (2006).
- Loubere, P., Mekik, F., Francois, R. & Pichat, S. Export fluxes of calcite in the eastern equatorial Pacific from the Last Glacial Maximum to present. *Paleoceanography* **19**, PA2018 (2004).
- Murray, R. W., Leinen, M. & Knowlton, C. W. Links between iron input and opal deposition in the Pleistocene equatorial Pacific Ocean. *Nature Geosci.* **5**, 270–274 (2012).
- Pichevin, L. E. *et al.* Enhanced carbon pump inferred from relaxation of nutrient limitation in the glacial ocean. *Nature* **459**, 1114–1117 (2009).
- Anderson, R. F. *et al.* Biological response to millennial variability of dust and nutrient supply in the Subantarctic South Atlantic Ocean. *Phil. Trans. R. Soc. Lond.* **372**, 20130054 (2014).
- Martinez-Garcia, A. *et al.* Iron fertilization of the subantarctic ocean during the last ice age. *Science* **343**, 1347–1350 (2014).
- Boyd, P. W. *et al.* Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science* **315**, 612–617 (2007).
- Hain, M. P., Sigman, D. M. & Haug, G. H. Carbon dioxide effects of Antarctic stratification, North Atlantic Intermediate Water formation, and subantarctic nutrient drawdown during the last ice age: diagnosis and synthesis in a geochemical box model. *Glob. Biogeochem. Cycles* **24**, 1–19 (2010).
- Rafter, P. A. & Sigman, D. M. Spatial distribution and temporal variation of nitrate nitrogen and oxygen isotopes in the upper equatorial Pacific Ocean. *Limnol. Oceanogr.* <http://dx.doi.org/10.1002/lno.10152> (2015).
- Kohfeld, K. E. & Harrison, S. P. DIRTMAP: the geological record of dust. *Earth Sci. Rev.* **54**, 81–114 (2001).
- McGee, D., Marcantonio, F. & Lynch-Stieglitz, J. Deglacial changes in dust flux in the eastern equatorial Pacific. *Earth Planet. Sci. Lett.* **257**, 215–230 (2007).
- Chase, Z., Anderson, R. F., Fleisher, M. Q. & Kubik, P. W. Scavenging of ^{230}Th , ^{231}Pa and ^{10}Be in the Southern Ocean (SW Pacific sector): the importance of particle flux and advection. *Deep Sea Res. Part II* **50**, 739–768 (2003).
- Hernandez-Sanchez, M. T. *et al.* Quantifying export production in the Southern Ocean: implications for the Ba_{xs} proxy. *Paleoceanography* **26**, PA4222 (2011).
- Robinson, R. S., Martinez, P., Pena, L. D. & Cacho, I. Nitrogen isotopic evidence for deglacial changes in nutrient supply in the eastern equatorial Pacific. *Paleoceanography* **24**, PA4213 (2009).
- Pichat, S. *et al.* Lower export production during glacial periods in the equatorial Pacific derived from ($^{231}\text{Pa}/^{230}\text{Th}$) $_{\text{xs},0}$ measurements in deep-sea sediments. *Paleoceanography* **19**, PA4023 (2004).
- Ren, H., Sigman, D. M., Thunell, R. C. & Prokopenko, M. G. Nitrogen isotopic composition of planktonic foraminifera from the modern ocean and recent sediments. *Limnol. Oceanogr.* **57**, 1011–1024 (2012).
- Altabet, M. A. & Francois, R. Sedimentary nitrogen isotopic ratio as a recorder for surface ocean nitrate utilization. *Glob. Biogeochem. Cycles* **8**, 103–116 (1994).
- Schroth, A. W., Crusius, J., Sholkovitz, E. R. & Bostick, B. C. Iron solubility driven by speciation in dust sources to the ocean. *Nature Geosci.* **2**, 337–340 (2009).
- Gordon, R. M., Coale, K. H. & Johnson, K. S. Iron distributions in the equatorial Pacific: implications for new production. *Limnol. Oceanogr.* **42**, 419–431 (1997).
- Labatut, M. *et al.* Iron sources and dissolved-particulate interactions in the seawater of the Western Equatorial Pacific, iron isotope perspectives. *Glob. Biogeochem. Cycles* **28**, 1044–1065 (2014).
- Reeves, J. M. *et al.* Palaeoenvironmental change in tropical Australasia over the last 30,000 years—a synthesis by the OZ-INTIMATE group. *Quat. Sci. Rev.* **74**, 97–114 (2013).
- Loubere, P. Marine control of biological production in the eastern equatorial Pacific Ocean. *Nature* **406**, 497–500 (2000).
- Rafter, P. A. & Charles, C. D. Pleistocene equatorial Pacific dynamics inferred from the zonal asymmetry in sedimentary nitrogen isotopes. *Paleoceanography* **27**, PA3102 (2012).
- Loubere, P. Nutrient and oceanographic changes in the Eastern Equatorial Pacific from the last full Glacial to the present. *Global Planet. Change* **29**, 77–98 (2001).
- Ford, H. L., Ravelo, A. C. & Polissar, P. J. Reduced El Niño–Southern Oscillation during the Last Glacial Maximum. *Science* **347**, 255–258 (2015).
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2004).
- Ho, T.-Y. *et al.* The elemental composition of some marine phytoplankton. *J. Phycol.* **39**, 1145–1159 (2003).
- Conkright, M. E. *et al.* in *NOAA Atlas NESDIS 51 Vol. 4 Nutrients* (ed. Levitus, S.) 1–392, <ftp://ftp.nodc.noaa.gov/pub/data.nodc/woa/publications/woa01v4d.pdf> (US Government Printing Office, 2002).

Acknowledgements We thank M. Soon at the University of British Columbia for assistance in running the opal analyses and M. A. Weigand and S. Oleynik for assistance with the $\text{FB-}\delta^{15}\text{N}$ analyses. Funding was provided in part by NSF award AGS 15-02889 (to J.F.McM., G.W. and F.M.), NSF award OCE-1060947 (to D.M.S.) and the Grand Challenges Program of Princeton University (to D.M.S.). This project also benefited from previous support from the NSF (grant numbers OCE-1003374, OCE-1159053, OCE-1158886) and the Comer Science and Education Foundation (to J.F.McM.) and travel funding from MOST, Taiwan (grant number 103-2116-M-002-032-MY2 to H.R.).

Author Contributions K.M.C., J.F.McM., and R.F.A. designed the study. K.M.C. performed the core sampling, U-Th-Pa chemistry, and instrumental analyses, with technical assistance from M.Q.F. Nitrogen isotope analyses were carried out by H.R. (for the Holocene) and K.M.C. with the assistance of H.R. (for the LGP). F.M. performed the barium measurements. All authors contributed to interpretation of the data. K.M.C. wrote the manuscript, and all authors provided comments and revisions.

Author Information All data presented in this manuscript have been deposited in the National Climatic Data Center (<https://www.ncdc.noaa.gov/paleo/study/19441>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M.C. (kcosta@ldeo.columbia.edu).

METHODS

Site location and chronology. The cores used in this study were collected aboard the RV *Marcus G. Langseth* on a dedicated cruise to the Line Islands in May 2012. The sites range from just south of the Equator (0.22 °S) to approximately 7° N along the 159° W ($\pm 3^\circ$) meridian (Extended Data Fig. 1, Extended Data Table 1). Shallow core sites (3,545 m at the deepest) suggest minimal carbonate dissolution, with carbonate concentrations ranging from 83–98 wt%. Core chronologies were established with four radiocarbon dates on *G. ruber*: 0 cm and 8 cm depth in multicores, and two depths in the Big Bertha piston cores that bracketed the $\delta^{18}\text{O}$ maximum inferred to represent Marine Isotope Stage 2 (ref. 31). Analyses were performed at the National Ocean Sciences Accelerator Mass Spectrometry Facility at Woods Hole Oceanographic Institution and at the Lawrence Livermore National Laboratory. Radiocarbon ages were calibrated to calendar years using Calib 7.0 Marine13^{32,33}. Age models were established via linear interpolation between radiocarbon dates (Extended Data Fig. 2, Extended Data Table 2). Core-top ages greater than zero are caused by bioturbation of the surficial sediment, resulting in inflated sedimentation rates in the multicores. Because our samples were selected based on $\delta^{18}\text{O}$ before radiocarbon dating, not all data points fall within the EPILOG definition of the Last Glacial Maximum (the peak of the LGP) of 18,000–24,000 years ago³⁴. However, amending the data to only those within that time frame ($n = 21$ of 31) has no substantial impact on the interpretations presented here.

Analytical methods. Approximately five samples from the multicores (Holocene) and five samples from the Big Bertha piston cores (LGP) from each of the six sites were analysed for thorium (^{230}Th , ^{232}Th), uranium (^{238}U , ^{235}U , ^{234}U), and protactinium (^{231}Pa) by isotope dilution inductively coupled plasma mass spectrometry (ICP-MS). Samples (100 mg) were randomized, spiked with ^{229}Th , ^{236}U and ^{233}Pa , and processed with complete acid digestion and column chromatography³⁵. Isotopes were measured on an Element 2 ICP-MS at the Lamont-Doherty Earth Observatory of Columbia University. An internal sediment standard (Line Islands MegaStandard, LIMS) was used for quality control, and replicates of LIMS ($n = 15$) indicate that measurements are reproducible within <5% on all isotopes. Excess initial ^{230}Th and excess initial ^{231}Pa values were calculated by correcting for supported decay (the fraction of ^{230}Th that is in equilibrium with its parent U isotope) from lithogenic and authigenic uranium³⁶. These corrections are negligible, with >98.9% of the total ^{230}Th scavenged from the water column and present as unsupported ^{230}Th (see section below on ^{230}Th normalization). Dust fluxes are calculated by normalizing the ^{232}Th concentration to ^{230}Th and using an average ^{232}Th concentration in equatorial Pacific dust of 10.7 parts per million¹³. Justification of ^{232}Th as a tracer for aeolian input is provided elsewhere³⁷.

Biogenic opal was measured by alkaline extraction³⁸ at the University of British Columbia (Extended Data Fig. 4). Opal fluxes were calculated by ^{230}Th normalization (see below). Opal fluxes reflect the burial of opal in the sediment, which is a function of both productivity and preservation, and so we compared the opal fluxes to $^{231}\text{Pa}/^{230}\text{Th}$, which is insensitive to preservation effects^{14,17}. Poor opal preservation may be a result of less silicified and thus more easily dissolvable diatom frustules, in response to the higher iron concentrations^{39–41} during the LGP, which would increase the $^{231}\text{Pa}/^{230}\text{Th}$ to opal ratios. Instead we observe no change in the $^{231}\text{Pa}/^{230}\text{Th}$ to opal ratio in the time slices presented here: $7.89 \pm 1.24 \times 10^{-3}$ in the Holocene, $8.09 \pm 1.31 \times 10^{-3}$ in the LGP, and $7.60 \pm 0.53 \times 10^{-3}$ overall (Extended Data Fig. 6). The temporal invariance of the $^{231}\text{Pa}/^{230}\text{Th}$ to opal ratio indicates both that there was no substantial change in diatom silicification due to iron stress and that the degree of opal preservation did not change substantially between the LGP and the Holocene⁴². Therefore we interpret changes in opal flux as changes in productivity.

Barium concentrations were analysed by total digestion at Texas A&M University of the same samples analysed for biogenic opal (Extended Data Figs 4 and 5). The standard deviation on barium concentration data is $\leq 4.3\%$, with an average of 2.8% for all samples. Excess barium concentrations were calculated by subtracting the lithogenic barium component, determined using a lithogenic Ba/Th mass ratio of 51.4 in the upper continental crust⁴³. Excess barium fluxes were calculated by ^{230}Th normalization (see below). Because dust fluxes are so low in the central equatorial Pacific, the lithogenic Ba fraction is small, <2% of the total Ba in the Holocene and <6% of the total Ba in the LGP. The lithogenic correction is insensitive to variability in the lithogenic Ba/Th ratio, so that variations over an order of magnitude (Ba/Th = 10–100) induce only small deviations (<10%) in the excess Ba fluxes (Extended Data Fig. 5). Excess barium in marine sediment is almost exclusively found in the form of barite, a robust sedimentary component insensitive to both dissolution and diagenesis¹⁵. The excess Ba flux may reflect the minimum amplitude for the change in productivity, as the proportional difference in excess Ba flux between the LGP and the Holocene

(up to 17%) is lower than that of $^{231}\text{Pa}/^{230}\text{Th}$ (up to 19%) and the opal flux (up to 33%) (Extended Data Fig. 7).

Foraminifera-bound nitrogen isotopes were analysed following the procedure outlined by ref. 18. Approximately 10 mg of *Globigerinoides ruber*, *Globigerinoides sacculifer*, and *Globobulimina tumida* were picked from five Holocene samples and two glacial samples from each of the six sites. Samples were chemically cleaned to remove nitrogen contamination, oxidized with persulfate to convert organic nitrogen to nitrate, measured by chemiluminescence for nitrate concentration, and converted to nitrous oxide by a bacterial denitrifier that lacks a functional nitrous oxide reductase. The $\delta^{15}\text{N}$ of nitrous oxide was measured by gas chromatography isotope mass spectrometry on a purpose-built nitrous oxide extraction and purification system on-line with a Thermo MAT 253 stable isotope ratio mass spectrometer at Princeton University. $\delta^{15}\text{N}$ is reported relative to atmospheric N_2 . Analyses of Holocene samples ($n = 5$ per core) by H.R. and LGP samples ($n = 2$ per core) by K.M.C. were performed in the same laboratory following the same procedure. The standard deviation for analysis of a picked and cleaned sample of a given foraminifera species is $\leq 0.3\%$, while the reproducibility for a given coarse fraction, including the species picking and cleaning, averages 0.3%, with a maximum of 0.8%.

^{230}Th normalization. The concentration of minor sedimentary components, such as opal and excess Ba, can vary in response both to changes in the input of the components themselves as well as to changes in the dilution by major sedimentary components, such as CaCO_3 . In the Pacific, carbonate is better preserved in glacial periods than in interglacial periods⁴⁴, and in the Line Islands we observed higher CaCO_3 concentrations (91%–93%) in the LGP than in the Holocene (86%–91%) (Extended Data Fig. 4). In the LGP we generally see lower opal and excess Ba concentrations, consistent with increased dilution by CaCO_3 . To remove the dilution effects and isolate changes in the opal and excess Ba inputs, we normalize to ^{230}Th , a constant flux proxy.

^{230}Th and ^{231}Pa are produced in the water column by the decay of ^{234}U and ^{235}U , respectively. Because U is highly soluble in sea water, it has a long residence time (400,000 years, or 400 kyr) and a fairly constant concentration (3.2 parts per billion) that scales conservatively with salinity⁴⁵. Thus the production of ^{230}Th and ^{231}Pa is relatively uniform across the global ocean and depends primarily on the water depth and to a minor extent on salinity. ^{230}Th is produced at a rate of 0.0262 disintegrations per minute (d.p.m.) $\text{m}^{-3} \text{yr}^{-1}$ (β_{230}), while ^{231}Pa is produced at a rate of 0.00245 d.p.m. $\text{m}^{-3} \text{yr}^{-1}$ (β_{231} , measured in activity units)⁴⁶. The production ratio of ^{231}Pa to ^{230}Th ($\beta_{230}/\beta_{231} = 0.093$) is assumed to be constant across the global ocean.

Unlike U, both Th and Pa are practically insoluble in sea water and thus have relatively short residence times in the ocean (20–40 years and 100–200 years, respectively)⁴⁷. ^{230}Th and ^{231}Pa are removed from sea water by scavenging onto settling particles, and they are buried in the underlying sediments unsupported by their parent nuclides. This excess ^{230}Th and ^{231}Pa in the sediment decays over time with respective half-lives of 75.69 kyr and 32 kyr (ref. 48).

The residence time of ^{230}Th in the ocean is so short (20–40 years) compared to its half-life that virtually all of the ^{230}Th produced by U decay in sea water is removed to sediments by scavenging. The residence time of ^{230}Th is also much less than the timescale for lateral transport by mixing from regions of low scavenging intensity (low particle flux) to regions of high scavenging intensity. Consequently, throughout the ocean the flux of ^{230}Th carried to the sea bed by sinking particles is within $\sim 30\%$ of its production rate in the overlying water column⁴⁶. Given a rate of supply that depends mainly on water depth, the concentration of excess ^{230}Th in the underlying sediment is a function of the sediment rain rate, or bulk mass flux (BMF). Higher BMF will dilute the excess ^{230}Th concentration in the sediment. Thus the BMF can be calculated by dividing the integrated ^{230}Th production in the overlying water column (β_z) by the excess initial ^{230}Th , the concentration of excess ^{230}Th in the sediment corrected for decay since deposition³⁶. ^{230}Th -normalized fluxes of specific sedimentary components i , such as opal and excess Ba, can be calculated as $f_i \times \text{BMF}$, where f_i is the fraction of component i in the sediment. Normalization to ^{230}Th removes the effects of dilution, and it can correct for lateral rather than vertical sediment inputs (such as focusing and winnowing).

The validity of the ^{230}Th -normalization technique has generated much debate in the palaeoceanographic community^{49–51}. ^{230}Th systematics have been called into question with regard to (1) dependence on age models and propagated uncertainties; (2) changes in production of ^{230}Th by uranium decay in the water column; (3) size fractionation of ^{230}Th as result of sediment focusing and winnowing; and (4) changes in the local rate of scavenging of ^{230}Th from the water column by sinking particles. Although these present important considerations when interpreting ^{230}Th -normalization, several lines of evidence nevertheless justify its usage in this study.

Anomalously low mass fluxes could be inferred to be a result of erroneously overcorrecting the radioactive decay of excess ^{230}Th in LGP sediments. This uncertainty becomes increasingly important as sediment ages, particularly after several half-lives of ^{230}Th (for example, >400 kyr). The relatively short timescales investigated here (15–20 kyr) are unlikely to be sensitive to small changes in the age models, as the age correction from excess ^{230}Th to excess initial ^{230}Th is only of the order of a 1.6%–30% increase in ^{230}Th concentration. For example, to equate the Holocene and LGP equatorial opal fluxes, holding all isotopic concentrations constant, the age of the LGP sample would need to be 2,000 years old rather than 21,400 years old. Such an extreme error in the age assignment can be ruled out because the cores presented here have internally consistent age models constrained by radiocarbon and by $\delta^{18}\text{O}$. Furthermore, evidence for systematic age offsets would be discernible in the $^{231}\text{Pa}/^{230}\text{Th}$ ratio. Because ^{231}Pa has a shorter half-life than ^{230}Th , the correction for radioactive decay is much more sensitive to age assignments, with the result that the sense of the error in the $^{231}\text{Pa}/^{230}\text{Th}$ ratio is opposite to the sense of the error in ^{230}Th -normalized flux. Increasing the mass flux (for example, by adjusting the sample age to be younger) will simultaneously decrease the $^{231}\text{Pa}/^{230}\text{Th}$ ratio. The presence of both lower $^{231}\text{Pa}/^{230}\text{Th}$ and lower ^{230}Th -normalized fluxes during the LGP indicates that the lower palaeoproductivity is not an artefact of inaccurate age corrections.

Changes in production of ^{230}Th by uranium decay in the water column are similarly unlikely. The residence time of dissolved uranium in the oceans is so long (200–400 kyr) that the production rate of ^{230}Th cannot change substantially during the 15–20-kyr interval between the LGP and Holocene.

Although sediment winnowing and focusing may influence the absolute ^{230}Th concentrations, it is improbable that they induce relative offsets between the Holocene and the LGP. First, it has been suggested that ^{230}Th -normalization of coarse carbonate may be unreliable, but recent evidence^{52,53} supports its utilization in the fine sediment fraction, which is compatible with our application of Th-normalization to dust, barite and opal. Second, focusing factors, which measure the relative contributions of lateral and vertical sediment inputs, are nearly identical within the time slices: 1.28 ± 0.27 in the LGP and 1.25 ± 0.32 in the Holocene. This consistency between the time slices makes it unlikely that the observed changes in dust flux and productivity are an artefact of the ^{230}Th systematics. Lastly, the slight sediment focusing (focusing factor > 1) at these sites mitigates the grain size and sorting effects on the isotopes that are associated with sediment redistribution⁵⁴ and particularly with winnowing⁵⁵. As pointed out in ref. 55, moderate variations in lateral focusing do not appear to be influential in modifying the $^{231}\text{Pa}/^{230}\text{Th}$ isotope ratios or their reflection of the local particle fluxes. The changes observed here are approximately 8% between time slices.

Finally, ^{230}Th -normalization assumes that all (100%) of the ^{230}Th produced in the water column is scavenged and buried in the sediment, but lower scavenging efficiencies (for example, 80%) would result in lower ^{230}Th concentrations and higher inferred mass fluxes. The only way to increase the scavenging efficiency of ^{230}Th is to increase the flux of particles responsible for scavenging thorium from the water column. The flux of ^{230}Th scavenged from the water column is quite insensitive to changes in particle flux⁵⁶. Although not rigorously calibrated, we might qualitatively estimate that a doubling of particle flux would lead to a 10% increase in flux of ^{230}Th . If the ^{230}Th -normalized mass flux were accurate initially, then the result of doubled particle flux would be to underestimate the true flux by 10%. Specifically, we would infer an 81% increase in flux, whereas the true change would be a 100% increase. Nevertheless, the reconstructed signal shows the direction of change accurately, even though the amplitude is lower than it should be. Consequently, when we derive a lower particle flux during the LGP using the ^{230}Th -normalization method, we can be confident of the direction of the change, and we know that if there is a bias in the derived result then the actual change was even greater than the one we infer.

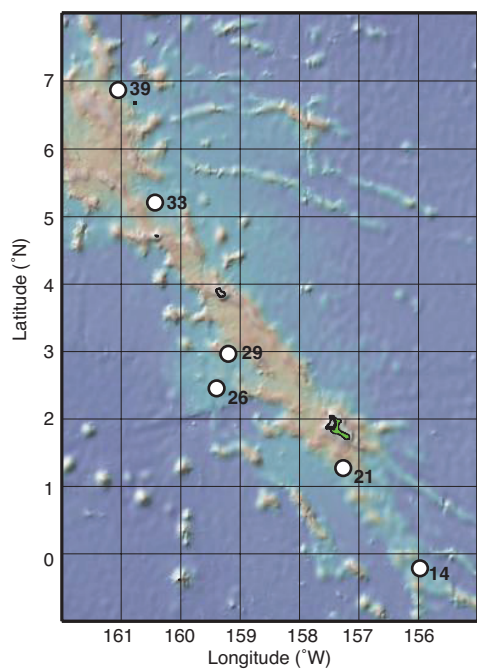
Denitrification and the $\delta^{15}\text{N}$ of upwelling waters. The eastern tropical Pacific contains two of the world's largest oxygen-deficient zones. Under these suboxic conditions, bioavailable nitrogen is lost via water column denitrification, which does not go to completion and thus imparts a strong fractionation to the nitrate pool, generating increased $\delta^{15}\text{N}$ in the residual nitrate. This high $\delta^{15}\text{N}$ nitrate is distributed from the Oxygen Minimum Zone throughout the Pacific Ocean by circulation and remineralization of organic matter⁵⁷. This $\delta^{15}\text{N}$ increase is recorded by organic matter deposited in the sediment, so that sedimentary $\delta^{15}\text{N}$ will tend to increase as a function of denitrification intensity. The influence of denitrification on sedimentary $\delta^{15}\text{N}$ is highest where partially consumed nitrate from an oxygen-deficient zone is upwelled into the surface, as occurs along the margin of the eastern tropical Pacific¹⁶; farther afield, the signal is diluted by lower $\delta^{15}\text{N}$ nitrate and, on a basin and global scale, countered by the input of low $\delta^{15}\text{N}$ from nitrogen fixation. The net result is that the $\delta^{15}\text{N}$ of nitrate in the Equatorial

Undercurrent, which upwells in the central equatorial Pacific, is largely unchanged from the initial $\delta^{15}\text{N}$ of nitrate in Subantarctic Mode Water⁵⁸. The Equatorial Undercurrent originates in the far western Pacific, and its eastward flow is rapid, which may explain why the $\delta^{15}\text{N}$ of its nitrate is so weakly affected by water column denitrification occurring in the eastern tropical Pacific.

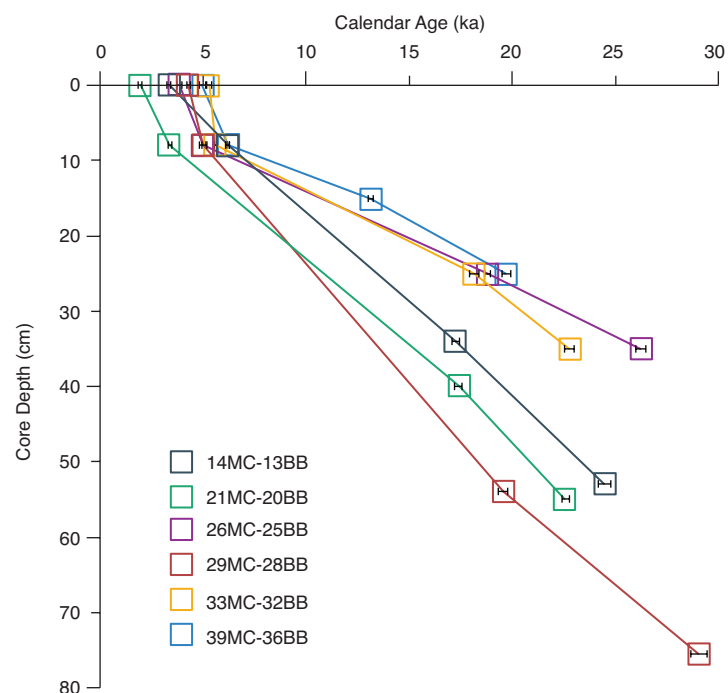
The intensity of water column denitrification has been reconstructed to be lower in the LGP than in the Holocene^{16,59–63}. With less elevation in the $\delta^{15}\text{N}$ of nitrate from the oxygen-deficient zones, bulk sedimentary $\delta^{15}\text{N}$ was lower by as much as 2.2‰ in the eastern tropical Pacific⁶³. However, the lower $\delta^{15}\text{N}$ does not appear to have propagated into the Equatorial Undercurrent. Records from the western equatorial Pacific, where bulk sedimentary $\delta^{15}\text{N}$ may reflect changes in the Equatorial Undercurrent source water, show little change: 0.03–0.14‰ lower in the LGP than in the Holocene²⁵. Thus, although we cannot eliminate the potential influence of reduced water column denitrification on the FB- $\delta^{15}\text{N}$ from the central equatorial Pacific (presented here), we infer that the effects are likely to be minimal. The north–south gradient in FB- $\delta^{15}\text{N}$ is insensitive to changes in the $\delta^{15}\text{N}$ of the nitrate upwelled along the Equator. Thus, the lack of change in this gradient between the Holocene and the LGP is further support that iron fertilization did not drive a substantial increase in nitrate consumption during the LGP, regardless of glacial–interglacial changes in denitrification.

- Lynch-Stieglitz, J. *et al.* Glacial-interglacial changes in central tropical Pacific surface seawater property gradients. *Paleoceanography* **30**, 423–438 (2015).
- Stuiver, M. & Reimer, P. J. Extended ^{14}C database and revised Calib 3.0 ^{14}C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
- Reimer, P. J. *et al.* Intcal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
- Mix, A. C., Bard, E. & Schneider, R. Environmental processes of the ice age: land, oceans, glaciers (EPILOG). *Quat. Sci. Rev.* **20**, 627–657 (2001).
- Fleisher, M. Q. & Anderson, R. F. Assessing the collection efficiency of Ross Sea sediment traps using ^{230}Th and ^{231}Pa . *Deep Sea Res. Part II* **50**, 693–712 (2003).
- Henderson, G. M. & Anderson, R. F. The U-series toolbox for paleoceanography. *Rev. Mineral. Geochem.* **52**, 493–531 (2003).
- Anderson, R., Fleisher, M. & Lao, Y. Glacial-interglacial variability in the delivery of dust to the central equatorial Pacific Ocean. *Earth Planet. Sci. Lett.* **242**, 406–414 (2006).
- Mortlock, R. A. & Froelich, P. N. A simple method for the rapid determination of biogenic opal in pelagic marine sediments. *Deep-Sea Res. Part A* **36**, 1415–1426 (1989).
- Takeda, S. Influence of iron availability on nutrient consumption ratio of diatoms in oceanic waters. *Nature* **393**, 774–777 (1998).
- Hutchins, D. A. & Bruland, K. W. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**, 561–564 (1998).
- Brzezinski, M. A. The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables. *J. Phycol.* **21**, 347–357 (1985).
- Warnock, J., Scherer, R. & Loubere, P. A quantitative assessment of diatom dissolution and late quaternary primary productivity in the Eastern Equatorial Pacific. *Deep Sea Res. Part II* **54**, 772–783 (2007).
- Taylor, S. R. & McLennan, S. M. The geochemical evolution of the continental crust. *Rev. Geophys.* **33**, 241–265 (1995).
- Farrell, J. W. & Prell, W. L. Climatic change and CaCO_3 preservation: an 800,000 year bathymetric reconstruction from the central equatorial Pacific Ocean. *Paleoceanography* **4**, 447–466 (1989).
- Owens, S. A., Buesseler, K. O. & Sims, K. W. W. Re-evaluating the ^{238}U -salinity relationship in seawater: implications for the ^{238}U - ^{234}Th disequilibrium method. *Mar. Chem.* **127**, 31–39 (2011).
- McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
- Hayes, C. T. *et al.* A new perspective on boundary scavenging in the North Pacific Ocean. *Earth Planet. Sci. Lett.* **369–370**, 86–97 (2013).
- Cheng, H. *et al.* The half-lives of uranium-234 and thorium-230. *Chem. Geol.* **169**, 17–33 (2000).
- Francois, R. *et al.* Comment on 'Do geochemical estimates of sediment focusing pass the sediment test in the equatorial Pacific?' by M. Lyle *et al.* *Paleoceanography* **22**, PA1216 (2007).
- Lyle, M., Mix, A. & Pisias, N. Patterns of CaCO_3 deposition in the eastern tropical Pacific Ocean for the last 150 kyr: evidence for a southeast Pacific depositional spike during marine isotope stage (MIS) 2. *Paleoceanography* **17**, 3–13–13 (2002).
- Lyle, M., Pisias, N., Paytan, A., Martinez, J. I. & Mix, A. Reply to comment by R. Francois *et al.* on 'Do geochemical estimates of sediment focusing pass the sediment test in the equatorial Pacific?': Further explorations of ^{230}Th normalization. *Paleoceanography* **22**, PA1217 (2007).
- Lyle, M. *et al.* Sediment size fractionation and focusing in the equatorial Pacific: effect on ^{230}Th normalization and paleoflux measurements. *Paleoceanography* **29**, 747–763 (2014).

53. Marcantonio, F., Lyle, M. & Ibrahim, R. Particle sorting during sediment redistribution processes and the effect on ^{230}Th -normalized mass accumulation rates. *Geophys. Res. Lett.* **41**, 5547–5554 (2014).
54. McGee, D., Broecker, W. S. & Winckler, G. Gustiness: the driver of glacial dustiness? *Quat. Sci. Rev.* **29**, 2340–2350 (2010).
55. Kretschmer, S. *et al.* Fractionation of ^{230}Th , ^{231}Pa , and ^{10}Be induced by particle size and composition within an opal-rich sediment of the Atlantic Southern Ocean. *Geochim. Cosmochim. Acta* **75**, 6971–6987 (2011).
56. Yu, E. F., Francois, R., Bacon, M. P. & Fleer, P. Fluxes of ^{230}Th and ^{231}Pa to the deep sea: implications for the interpretation of excess ^{230}Th and $^{231}\text{Pa}/^{230}\text{Th}$ profiles in sediments. *Earth Planet. Sci. Lett.* **191**, 219–230 (2001).
57. Sigman, D. M., DiFiore, P. J., Hain, M. P., Deutsch, C. & Karl, D. M. Sinking organic matter spreads the nitrogen isotope signal of pelagic denitrification in the North Pacific. *Geophys. Res. Lett.* **36**, L08605 (2009).
58. Rafter, P. A., Sigman, D. M., Charles, C. D., Kaiser, J. & Haug, G. H. Subsurface tropical Pacific nitrogen isotopic composition of nitrate: biogeochemical signals and their transport. *Glob. Biogeochem. Cycles* **26**, GB1003 (2012).
59. Martinez, P. & Robinson, R. S. Increase in water column denitrification during the last deglaciation: the influence of oxygen demand in the eastern equatorial Pacific. *Biogeosciences* **7**, 1–9 (2010).
60. Altabet, M., Francois, R., Murray, D. E. & Prell, W. L. Climate-related variations in denitrification in the Arabian Sea from sediment $^{15}\text{N}/^{14}\text{N}$ ratios. *Nature* **373**, 506–509 (1995).
61. De Pol-Holz, R. *et al.* Melting of the Patagonian ice sheet and deglacial perturbations of the nitrogen cycle in the eastern South Pacific. *Geophys. Res. Lett.* **33**, L04704 (2006).
62. Ganeshram, R. J., Pedersen, T., Calvert, S. E., McNeill, G. W. & Fontugne, M. R. Glacial–interglacial variability in denitrification in the world’s oceans: causes and consequences. *Paleoceanography* **15**, 361–376 (2000).
63. Dubois, N. *et al.* Millennial-scale variations in hydrography and biogeochemistry in the eastern equatorial Pacific over the last 100 kyr. *Quat. Sci. Rev.* **30**, 210–223 (2011).
64. Ryan, W. B. F. *et al.* Global multi-resolution topography synthesis. *Geochem. Geophys. Geosyst.* **10**, Q03014 (2009).

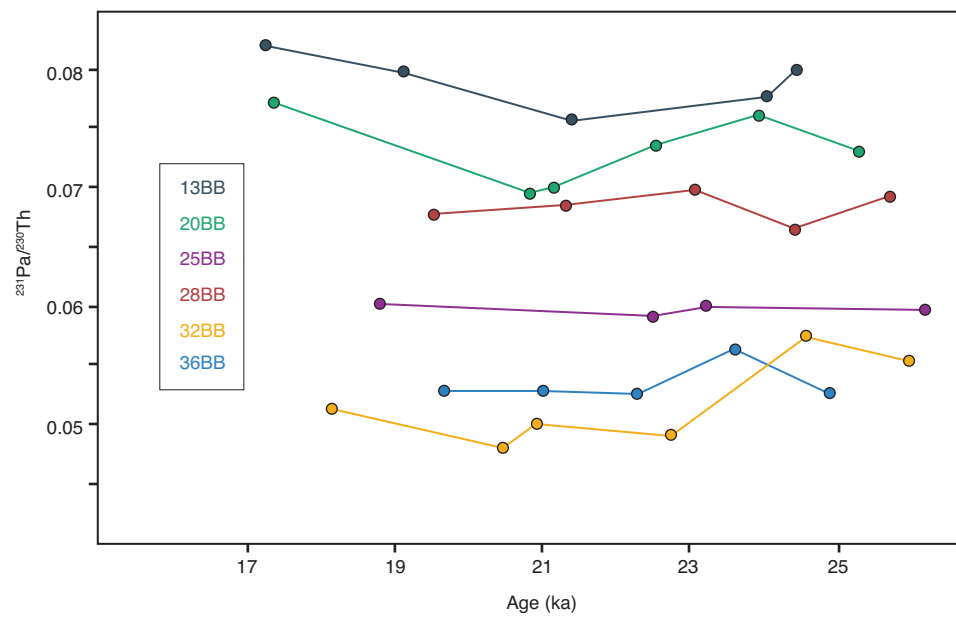


Extended Data Figure 1 | Line Islands core locations. Core sites are identified by their multicore numbers. The respective piston core numbers as well as latitude, longitude and depth are provided in Extended Data Table 1. The bathymetric map was generated using GeoMapApp⁶⁴.

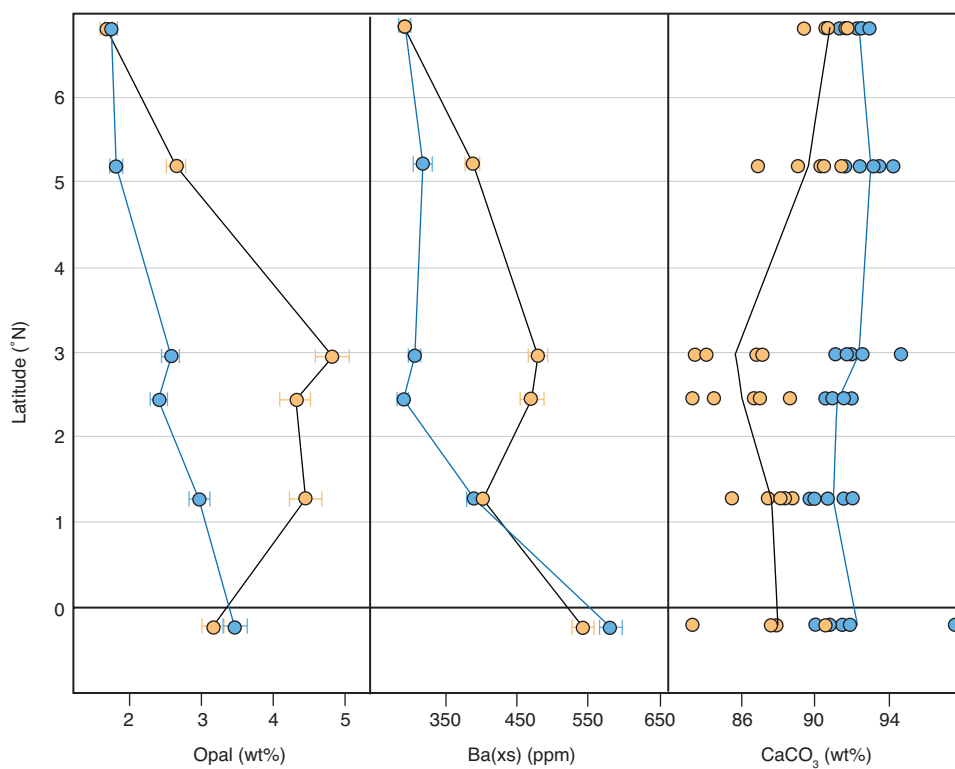


Extended Data Figure 2 | Radiocarbon-based age models for all six cores. Core chronologies were established with four radiocarbon dates on *G. ruber*: 0 cm and 8 cm depth in the multicores (MC), and two depths (>8 cm) in the Big Bertha piston cores (BB) bracketing the $\delta^{18}\text{O}$ maximum

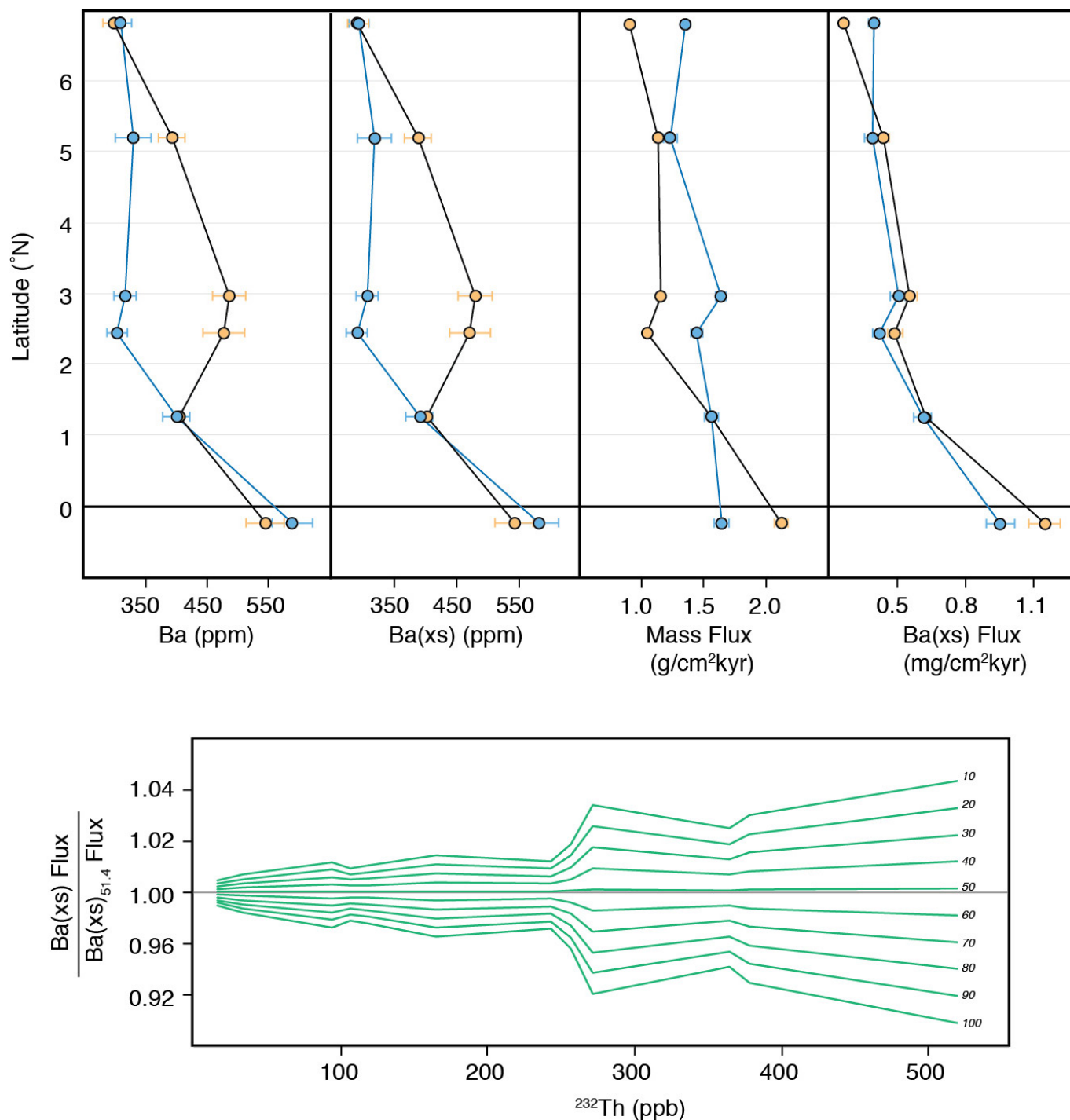
inferred to represent Marine Isotope Stage 2 (ref. 31). Age models were established via linear interpolation between radiocarbon dates (ka, thousands of years ago), which are provided in Extended Data Table 2.



Extended Data Figure 3 | Time series for $^{231}\text{Pa}/^{230}\text{Th}$ data within the LGP time slice. The relatively constant values for each core argue against any systematic bias from bioturbation of transient features, such as a deglacial productivity peak.

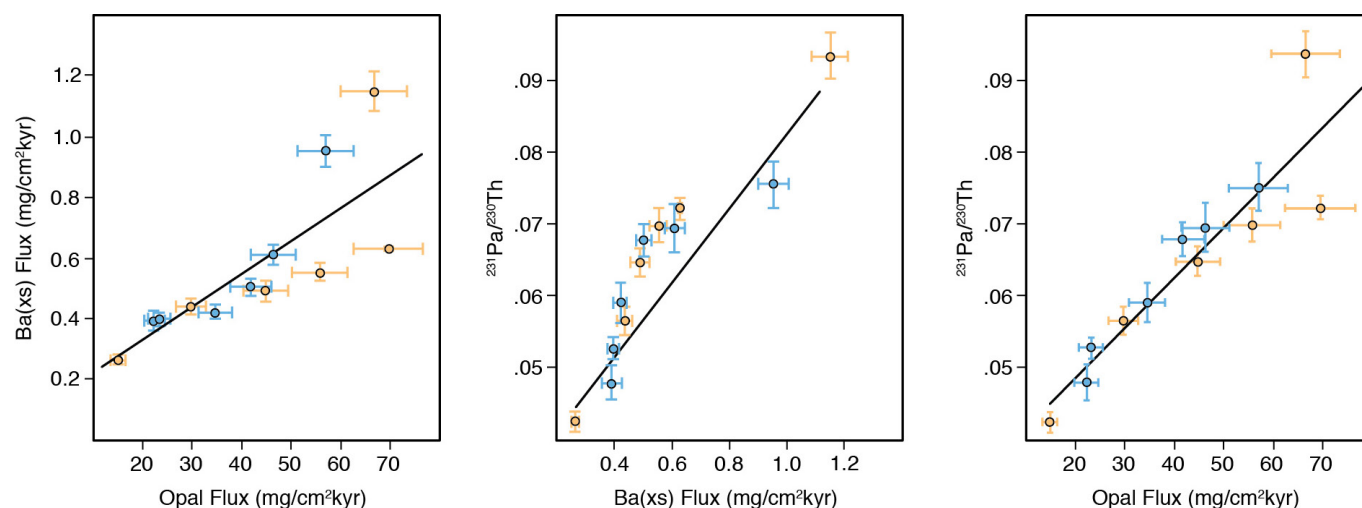


Extended Data Figure 4 | Concentrations of opal, excess Ba and CaCO_3 . CaCO_3 is the dominant sedimentary component, and systematic changes in its concentration dilute the concentrations of minor sedimentary components such as opal and excess Ba (Ba_{xs}).



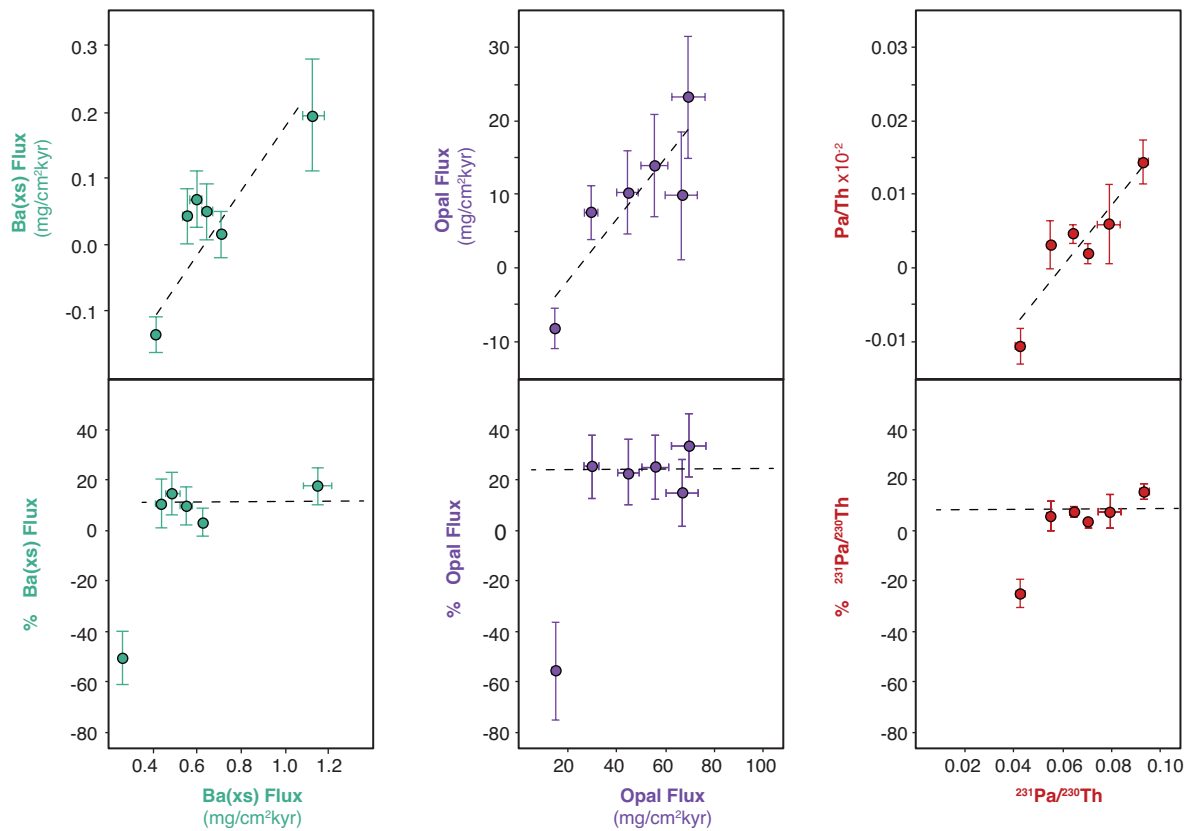
Extended Data Figure 5 | Lithogenic correction for Ba excess flux calculation. Samples from the Holocene (0–10,000 years ago) are orange; samples from the LGP (17,000–27,000 years ago) are blue. In the top panel, excess Ba concentrations were calculated by subtracting the lithogenic Ba fraction from the total Ba concentration using a lithogenic Ba/Th ratio of 51.4, based on the average elemental concentrations in upper continental crust⁴³. The lithogenic corrections are small, <2% for the Holocene and <6% for the LGP. Excess Ba concentrations were then multiplied by the

total mass flux (²³⁰Th-normalized) in order to generate the excess Ba flux. The bottom panel shows a comparison of excess Ba fluxes calculated using different lithogenic Ba/Th ratios, ranging from 10 to 100, normalized to the fluxes determined using a Ba/Th ratio of 51.4. Ratios less than 51.4 result in slightly higher excess Ba fluxes, while ratios greater than 51.4 result in slightly lower excess Ba fluxes. Overall, the excess Ba fluxes are insensitive to the Ba/Th ratio chosen, with deviations only over a range of ±10%. ppm, parts per million; ppb, parts per billion.



Extended Data Figure 6 | ^{230}Th -normalized opal flux, excess Ba flux, and $^{231}\text{Pa}/^{230}\text{Th}$. Samples from the Holocene (0–10,000 years ago) are orange; samples from the LGP (17,000–27,000 years ago) are blue. $^{231}\text{Pa}/^{230}\text{Th}$ is positively correlated with the opal flux ($r^2 = 0.90$, $P < 0.001$) and excess Ba flux ($r^2 = 0.85$, $P < 0.01$). Excess Ba flux and opal flux are also positively correlated ($r^2 = 0.63$, $P < 0.01$). For opal flux, the correlation with $^{231}\text{Pa}/^{230}\text{Th}$ is especially strong during the LGP (opal flux $r^2 = 0.98$, $P < 0.001$). The relationship (that is, the slope) between $^{231}\text{Pa}/^{230}\text{Th}$ and opal flux may be altered by changes in preservation, which affects opal

but not $^{231}\text{Pa}/^{230}\text{Th}$. Poor opal preservation (for example, if the diatom frustules were less silicified) would elevate the $^{231}\text{Pa}/^{230}\text{Th}$ relative to the sedimentary opal flux, thus steepening the slope. However, the relationship between $^{231}\text{Pa}/^{230}\text{Th}$ and opal flux is temporally invariant, with slopes of $7.89 \pm 1.24 \times 10^{-3}$ in the Holocene, $8.09 \pm 1.31 \times 10^{-3}$ in the LGP, and $7.60 \pm 0.53 \times 10^{-3}$ overall. We interpret these results to indicate that there was no important change in opal preservation between the LGP and Holocene and, therefore, that frustule silicification (potentially related to iron stress^{32–34}), similarly remained unchanged.



Extended Data Figure 7 | Absolute and relative change for each productivity proxy versus the Holocene value of that proxy. The absolute change (Holocene minus LGP) in productivity proxy is shown in the top panels; the relative change is shown in the bottom panels. Excess Ba flux is shown in green, opal flux in purple, and $^{231}\text{Pa}/^{230}\text{Th}$ in red. The greatest change in productivity occurred at the sites with the highest Holocene productivity values. The core that shows a negative change in productivity is the most northerly core (7° N), which is outside the high-nutrient, low-chlorophyll equatorial upwelling zone, and thus displays different

glacial–interglacial nutrient dynamics. The relative change in productivity (Holocene to the LGP) is fairly constant across the five cores within the equatorial upwelling zone at 11% for excess Ba flux, 24% for opal flux, and 8% for $^{231}\text{Pa}/^{230}\text{Th}$. The inter-proxy difference may reflect a nonlinear scaling of productivity with $^{231}\text{Pa}/^{230}\text{Th}$, because these radionuclides are scavenged to some extent by all particle phases. In practice, this difference suggests that $^{231}\text{Pa}/^{230}\text{Th}$ may provide a conservative estimate for changes in productivity, with true productivity changes potentially at much higher amplitude. Error bars are 2σ and indicate analytical precision.

Extended Data Table 1 | Locations of the Line Islands cores

MC #	BB #	Lat (°N)	Long (°W)	Depth (m)
14	13	-0.22	156.0	3049
21	20	1.27	157.3	2850
26	25	2.46	159.4	3545
29	28	2.97	159.2	3152
33	32	5.20	160.4	2933
39	36	6.83	161.0	2859

Extended Data Table 2 | Radiocarbon ages for age model generation

Core	Depth (cm)	Radiocarbon Age (yrs)	±	Calendar Age (yrs)	±
14MC	0	3460	35	3340	110
14MC	8	5770	30	6200	90
13BB	34	14580	40	17260	180
13BB	53	20760	70	24440	250
21MC	0	2330	25	1950	80
21MC	8	3500	35	3390	110
20BB	40	14685	40	17400	200
20BB	55	19100	70	22540	160
26MC	0	3900	35	3880	130
26MC	8	4740	35	4980	130
25BB	25	15950	60	18800	140
25BB	35	22400	100	26200	230
29MC	0	4200	15	4300	80
29MC	8	4770	20	5020	110
28BB	54	16600	60	19550	220
28BB	76	25400	160	29030	360
33MC	0	4930	35	5270	150
33MC	8	5260	20	5610	40
32BB	25	15350	75	18170	220
32BB	35	19300	75	22750	240
39MC	0	4730	45	4960	130
39MC	8	5800	20	6230	80
36BB	15	11650	25	13150	140
36BB	25	16700	90	19670	190

Error bars are 2σ and indicate analytical precision

Proton-gated Ca^{2+} -permeable TRP channels damage myelin in conditions mimicking ischaemia

Nicola B. Hamilton¹, Karolina Kolodziejczyk^{1†}, Eleni Kougioumtzidou¹ & David Attwell¹

The myelin sheaths wrapped around axons by oligodendrocytes are crucial for brain function. In ischaemia myelin is damaged in a Ca^{2+} -dependent manner, abolishing action potential propagation^{1,2}. This has been attributed to glutamate release activating Ca^{2+} -permeable *N*-methyl-D-aspartate (NMDA) receptors^{2–4}. Surprisingly, we now show that NMDA does not raise the intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) in mature oligodendrocytes and that, although ischaemia evokes a glutamate-triggered membrane current⁴, this is generated by a rise of extracellular $[\text{K}^+]$ and decrease of membrane K^+ conductance. Nevertheless, ischaemia raises oligodendrocyte $[\text{Ca}^{2+}]_i$, $[\text{Mg}^{2+}]_i$ and $[\text{H}^+]_i$, and buffering intracellular pH reduces the $[\text{Ca}^{2+}]_i$ and $[\text{Mg}^{2+}]_i$ increases, showing that these are evoked by the rise of $[\text{H}^+]_i$. The H^+ -gated $[\text{Ca}^{2+}]_i$ elevation is mediated by channels with characteristics of TRPA1, being inhibited by ruthenium red, isopentenyl pyrophosphate, HC-030031, A967079 or TRPA1 knockout. TRPA1 block reduces myelin damage in ischaemia. These data suggest that TRPA1-containing ion channels could be a therapeutic target in white matter ischaemia.

Ischaemia blocks action potential propagation through myelinated axons¹. Electron microscopy² and imaging of dye-filled oligodendrocytes³ show ischaemia-evoked Ca^{2+} -dependent damage to the capacitance-reducing myelin sheaths, which causes loss of action potential propagation. Glutamate receptor block reduces myelin damage and action potential loss^{2–7}, and glutamate evokes a membrane current in oligodendrocytes mediated by AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid)/kainate and NMDA receptors^{2–4}. Thus, oligodendrocyte damage is thought to be excitotoxic: as for neurons in ischaemia, a rise of glutamate concentration⁸ caused by reversal of glutamate transporters in oligodendrocytes and axons^{9,10} activates receptors that raise² oligodendrocyte $[\text{Ca}^{2+}]_i$, thus damaging the cells.

However, although AMPA/kainate and NMDA receptors regulate oligodendrocyte precursor development^{11,12}, these receptors are down-regulated as the cells mature^{13–15}. How can mature oligodendrocytes be damaged excitotoxically, if they express low levels of glutamate receptors? To investigate how oligodendrocyte $[\text{Ca}^{2+}]_i$ is raised in ischaemia, we characterized ischaemia-evoked membrane current and $[\text{Ca}^{2+}]_i$ changes in cerebellar white matter oligodendrocytes.

Solution mimicking ischaemia (see Methods) evoked an increasing inward current in oligodendrocytes (Fig. 1a, b), often with a faster phase that was obscured when responses in many cells were averaged (Fig. 1c). When applied from before the ischaemia, NBQX and D-AP5 reduced the ischaemia-evoked current by 66% (Fig. 1c, d), while mGluR block had no effect (Extended Data Fig. 1a). Preloading for 30 min with the glutamate transport blocker PDC, to prevent ischaemia-evoked glutamate release by reversal of transporters in the white⁹ and grey¹⁶ matter, also reduced the inward current (by 68%, Fig. 1c, d), while blocking other candidate release mechanisms had no effect (Extended Data Fig. 1a). Thus, glutamate release by reversed uptake helps to trigger the ischaemia-evoked current. Notably, however, current flow through glutamate receptors generates only a small fraction of the sustained inward

current evoked by ischaemia, since applying NBQX and D-AP5 from 200 s after ischaemia had started produced only a non-significant 21% suppression of the ischaemia-evoked inward current (Fig. 1d).

In neurons, an ischaemia-evoked inward current triggered by glutamate release, but maintained by non-glutamatergic mechanisms, generates the ‘extended neuronal depolarization’ (END) that evokes neuronal death¹⁷. However, the ischaemia-evoked current in oligodendrocytes was not prevented by removing external Ca^{2+} , nor by gadolinium, which both block the END¹⁷ (Fig. 1d, e), implying a different mechanism maintains the inward current triggered by glutamate.

Unlike in neurons, where ischaemia evokes a conductance increase mediated by ionotropic glutamate receptors¹⁶, ischaemia decreased the conductance of oligodendrocytes (Fig. 1f, g). The suppressed current reversed below the K^+ reversal potential ($E_K = -104$ mV), at -121 mV with 10 mM (Fig. 1f, Extended Data Fig. 1e) and -118 mV with 0.5 mM HEPES (Fig. 1g). This is expected if ischaemia decreases the membrane K^+ conductance, while $[\text{K}^+]_o$ rises (due to Na^+/K^+ pump inhibition throughout the slice) which increases the inward current at all potentials (see Extended Data Fig. 1b, d; Supplementary Information). K^+ -sensitive electrodes showed that $[\text{K}^+]_o$ in the white matter initially rose slowly during ischaemia, but then increased more abruptly, in parallel with the membrane current (Fig. 1h). The peak rise was 2.35 ± 0.13 mM ($n = 12$) in the white matter and 2.48 ± 0.35 mM ($n = 4$) in the adjacent grey matter (where it reflects the anoxic depolarization of neurons¹⁸). The conductance decrease described above produced 32%, while the $[\text{K}^+]_o$ rise produced 68%, of the inward current in oligodendrocytes at -74 mV (see Supplementary Information). Thus, changes in K^+ fluxes generate the ischaemia-evoked inward current.

Could part of the NMDA-evoked inward current in oligodendrocytes⁴ also reflect a $[\text{K}^+]_o$ rise? Extracellular Cs^+ blocked the NMDA-evoked current while intracellular MK-801 had no effect (Extended Data Fig. 2a–d), suggesting that most of the NMDA-evoked current is generated by $[\text{K}^+]_o$ rising, rather than by oligodendrocyte NMDA receptors. Applying NMDA or raising $[\text{K}^+]_o$, and correlating the resulting inward current with the $[\text{K}^+]_o$ rise occurring (see Supplementary Information; Extended Data Fig. 2e, f), we found that at least 49% of the NMDA-evoked current was attributable to the $[\text{K}^+]_o$ rise that it produced. Since mature oligodendrocytes express few NMDA receptors^{13–15}, this presumably reflects NMDA depolarizing neurons or astrocytes in the slice and releasing K^+ .

These data challenge the idea^{2–5} that, during ischaemia, NMDA receptors in mature oligodendrocytes generate a prolonged calcium influx which damages the cells. We therefore investigated the ion concentration changes evoked in oligodendrocytes by activation of NMDA receptors, using Ca^{2+} -, Na^+ - and K^+ -sensitive dyes loaded into cells from the pipette. When 100 μM NMDA was applied to whole-cell clamped cerebellar granule neurons at -74 mV, as expected it evoked an inward current, and raised $[\text{Ca}^{2+}]_i$ and $[\text{Na}^+]_i$ (Extended Data Fig. 3). In contrast, although NMDA evoked an inward current in oligodendrocytes, it generated no $[\text{Ca}^{2+}]_i$ or $[\text{Na}^+]_i$ elevation; indeed $[\text{Na}^+]_i$

¹Department of Neuroscience, Physiology & Pharmacology, University College London, Gower St., London WC1E 6BT, UK. [†]Present address: Department of Psychiatry, University of British Columbia, 2255 Wesbrook Mall, Vancouver, British Columbia V6T 1Z3, Canada.

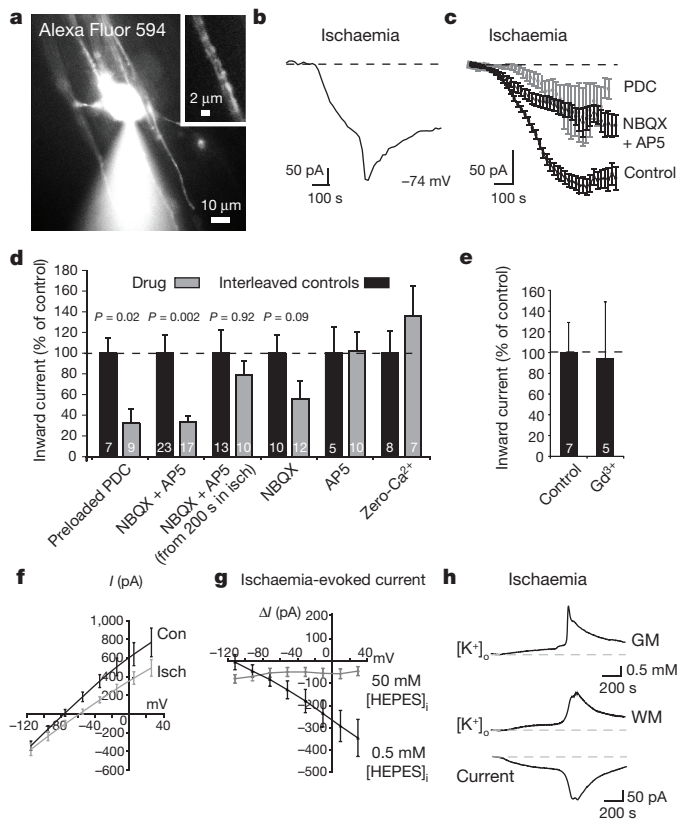


Figure 1 | Ischaemia evokes an inward current in oligodendrocytes by altering K^+ fluxes. **a**, Whole-cell clamped rat oligodendrocyte. Inset, Alexa dye in processes around an axon. **b**, Ischaemia-evoked membrane current in single cell. **c**, Current in 179 control cells, 12 cells in slices exposed to 25 μ M NBQX and 200 μ M D-AP5 from before ischaemia, or 9 cells in slices preloaded¹⁶ with 1 mM PDC. **d**, Current (normalized to interleaved control cells) from 8–10 min after start of ischaemia in cells preloaded with PDC, exposed to NBQX+AP5 throughout ischaemia or from 200 s after ischaemia starts, or exposed to NBQX or AP5 alone or to zero-Ca²⁺ solution (with 50 μ M EGTA) throughout ischaemia. Mann–Whitney P values compare with control cells; cell numbers shown on bars. **e**, Effect of Gd³⁺ (100 μ M) on ischaemia-evoked current at 8–10 min (Mann–Whitney $P = 0.83$). **f**, I – V relation of 10 cells before and after 5 min ischaemia (10 mM HEPES internal). **g**, Ischaemia-evoked current in 10 cells with 0.5 mM and 9 cells with 50 mM internal HEPES. Ischaemia decreased cell conductance by 2.1 ± 0.7 nS near -70 mV in 11 cells using 10 mM, and by 2.3 ± 0.6 nS in 10 cells using 0.5 mM, internal HEPES; 50 mM HEPES abolished the decrease (Fig. 3i). **h**, Change of $[K^+]_o$ in grey matter (GM, granule cell layer), and in white matter (WM, different slice) with simultaneously recorded oligodendrocyte current. Error bars, s.e.m.

decreased after applying NMDA (Fig. 2a). NMDA raised $[K^+]_i$ however. Similar concentration changes were seen at the soma (Fig. 2a) and in the internodal processes where NMDA receptors may be located^{2–4} (Fig. 2b). Like NMDA, raising $[K^+]_o$ lowered $[Na^+]_i$ (Fig. 2a, b). A likely explanation is that NMDA raises $[K^+]_o$ (Fig. 2c), which decreases $[Na^+]_i$ by activating the Na^+/K^+ pump.

The absence of a rise of $[Ca^{2+}]_i$ and $[Na^+]_i$ is surprising if oligodendrocytes express NMDA receptors^{2–4}. Conceivably, NMDA receptors might pass ions into a compartment in their myelinating processes which only certain Ca^{2+} -sensing dyes such as X-Rhod-1 can access². However, whether X-Rhod-1 was loaded as an acetoxymethyl ester² or from the pipette, we observed no NMDA-evoked change of $[Ca^{2+}]_i$ in the myelinating processes (Fig. 2d). Nevertheless, we could detect spontaneous $[Ca^{2+}]_i$ rises propagating through myelinating processes in 55% of oligodendrocytes (Fig. 2e).

To confirm that ischaemia raises² oligodendrocyte $[Ca^{2+}]_i$, we loaded the Ca^{2+} -sensing dye Fluo-4 (with Alexa Fluor 594, for ratiometric

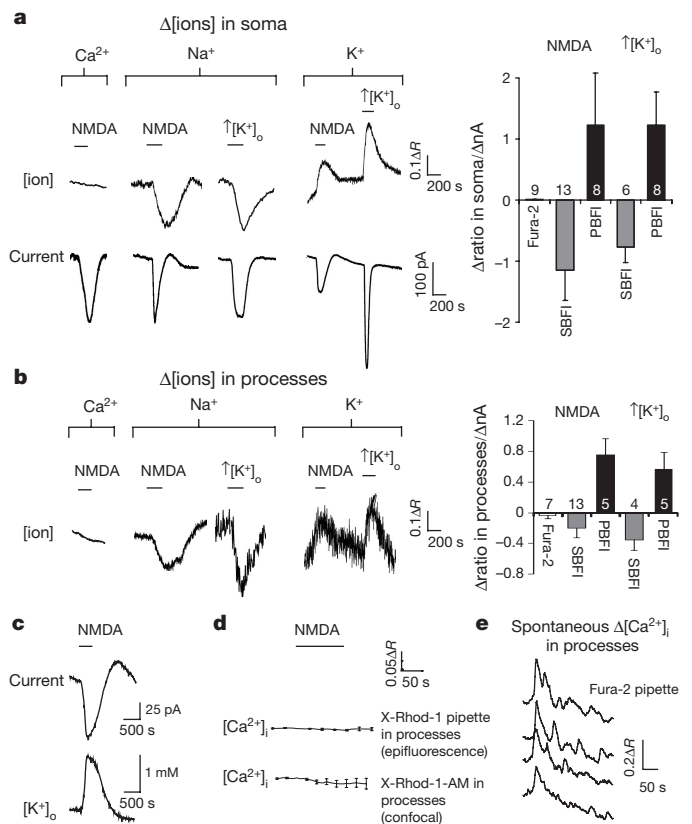


Figure 2 | NMDA does not elevate $[Ca^{2+}]_i$ in oligodendrocytes.

a, b, Rat oligodendrocyte membrane current (lower traces, **a**) and background-subtracted fluorescent dye ratio (R , see Methods, concentration increases are upwards for all dyes) when measuring $[Ca^{2+}]_i$ with Fura-2, $[Na^+]_i$ with SBFI, and $[K^+]_i$ with PBFI; 100 μ M NMDA was applied, or $[K^+]_o$ was raised from 2.5 to 5 mM, with fluorescence measured in soma (**a**) or myelinating processes (**b**). Right panels, peak fluorescence change normalized to evoked current (number of cells on bars) **c**, NMDA-evoked current and simultaneously recorded $[K^+]_o$. **d**, Measuring $[Ca^{2+}]_i$ with X-Rhod-1, loaded from pipette ($n = 6$) or as an acetoxymethyl ester² ($n = 15$), reveals no NMDA-evoked $[Ca^{2+}]_i$ rise. **e**, Spontaneous $[Ca^{2+}]_i$ transients in four myelinating processes confirm Fura-2 is working. Error bars, s.e.m.

imaging) into oligodendrocytes from a whole-cell pipette (in current-clamp mode, allowing voltage changes, in case voltage-gated Ca^{2+} channels raise $[Ca^{2+}]_i$). Ischaemia increased $[Ca^{2+}]_i$ in the soma and processes over ~ 10 min. This was abolished if extracellular calcium was removed (Fig. 3a, b), and reduced by removing external K^+ (Fig. 3c), suggesting that the ischaemia-evoked $[K^+]_o$ rise promotes calcium entry from the extracellular solution. However, contradicting the earlier report², blocking NMDA receptors with MK-801, D-AP5 and 7-chloro-kynurenat, or blocking NMDA and AMPA/kainate receptors with NBQX and D-AP5 while blocking voltage-gated Na^+ and Ca^{2+} channels and GABA_A receptors, did not prevent the $[Ca^{2+}]_i$ rise (Fig. 3a, b). Similarly, when PDC-preloading reduced transporter-mediated glutamate release, the $[Ca^{2+}]_i$ rise was unaffected (Fig. 3c).

Similar experiments using a Mg^{2+} -sensitive dye revealed that $[Mg^{2+}]_i$ also rises in ischaemia (Fig. 3d). This was not due to ATP breakdown, which releases Mg^{2+} , since the $[Mg^{2+}]_i$ rise was abolished by removing extracellular Mg^{2+} (Fig. 3d) implying that Mg^{2+} enters across the cell membrane. Surprisingly, ischaemia did not raise $[Na^+]_i$ (Fig. 3e). Thus ischaemia activates a membrane conductance that allows entry of divalent ions.

Seeking an agent that decreases membrane K^+ conductance and activates Ca^{2+} entry, we measured the ischaemia-evoked pH change in oligodendrocytes. Ischaemia increased $[H^+]_i$ on the timescale seen for $[Ca^{2+}]_i$ (Fig. 3f). A similar (but smaller) $[H^+]_i$ rise was evoked by

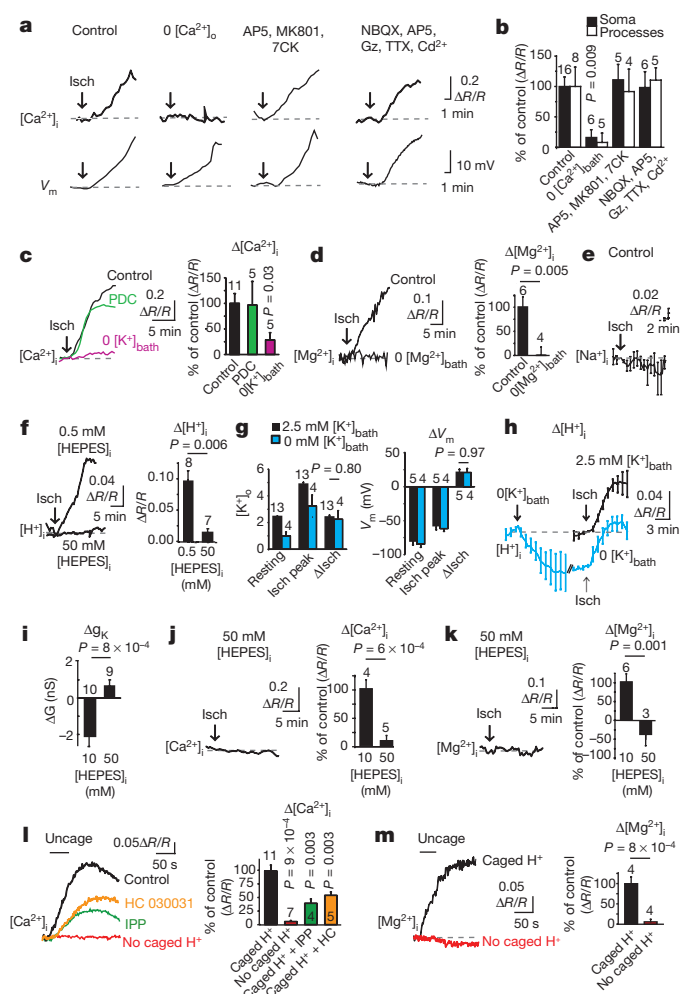


Figure 3 | Ischaemia evokes a [Ca²⁺]_i and [Mg²⁺]_i rise gated by internal protons. **a**, Ratiometric Fluo-4/Alexa-Fluor-594 signals ([Ca²⁺]_i) and membrane potential (V_m) in rat oligodendrocytes exposed to ischaemia (starting at arrow), or ischaemia in zero [Ca²⁺]_o, or with drugs at concentrations (μM): AP5 50, MK-801 50, 7-chlorokynurenate (7CK) 100, NBQX 25, GABAzine (Gz) 20, TTX 1, Cd²⁺ (to block Ca²⁺ channels) 100. **b**, Mean data from experiments like **a** (cell numbers on bars; *P* values compare with soma or process control values). **c**, Data as in **a** after PDC-preloading or in 0 mM [K⁺]_{bath}. **d**, Ischaemia-evoked [Mg²⁺]_i rise monitored with Mag-Fluo-4 in normal and Mg²⁺-free solution. **e**, Ischaemia-evoked [Na⁺]_i change monitored with SBFI (6 cells). **f**, Ischaemia-evoked [H⁺]_i rise monitored with BCECF with 0.5 mM and 50 mM internal HEPES (*P* from Mann–Whitney test). **g**, [K⁺]_o and oligodendrocyte membrane potential (V_m) with 2.5 or 0 mM [K⁺]_{bath}, before (Resting) and during ischaemia (Isch peak), and change produced by ischaemia (ΔIsch). **h**, Effect of removing K⁺ from bath solution on [H⁺]_i in control conditions (relative to value at start of K⁺ removal), and [H⁺]_i increase evoked by ischaemia (normalized to value at start of ischaemia) with 2.5 or 0 mM bath K⁺. **i**, High [HEPES]_i blocks ischaemia-evoked decrease of membrane conductance. **j**, **k**, Ischaemia-evoked rise of [Ca²⁺]_i (**j**) and [Mg²⁺]_i (**k**) are inhibited with 50 mM internal HEPES. **l**, **m**, Uncaging H⁺ with light (bars) raises [Ca²⁺]_i (**l**, an effect reduced by 200 μM IPP or 80 μM HC-030031) and [Mg²⁺]_i (**m**), but not when caged H⁺ is omitted from the pipette. *P* values in **l** from Mann–Whitney tests. Error bars, s.e.m.

elevated [K⁺]_o or NMDA (Extended Data Fig. 4a), suggesting that the ischaemia-evoked [K⁺]_o rise partly generates this pH change. To investigate this, we removed external K⁺, which reduced the [K⁺]_o in the slice from 2.46 ± 0.02 mM (*n* = 13) to 0.99 ± 0.30 mM (*n* = 4) (Mann–Whitney *P* = 0.001, Fig. 3g), and hyperpolarized the resting potential by 7 mV (−84.0 ± 4.7 to −77.2 ± 4.3 mV, *n* = 4, *P* = 0.002, paired *t*-test, Fig. 3g). The ischaemia-evoked [K⁺]_o rise and depolarization were

unaffected by K⁺ removal (Fig. 3g), but the [H⁺]_i initially and during ischaemia were reduced (Fig. 3h). These data and those in Fig. 3c suggest that the ischaemia-evoked [K⁺]_o rise helps to acidify the cell, which in turn evokes Ca²⁺ entry.

Using an internal solution containing 50 mM HEPES, the ischaemia-evoked rise of [H⁺]_i was, as expected, greatly reduced (Fig. 3f). This prevented the decrease of K⁺ conductance (Figs 1g, 3i), consistent with intracellular acidity suppressing activity of tonically-active K⁺ channels¹⁹. Strikingly, however, buffering intracellular pH also prevented the ischaemia-evoked rise of [Ca²⁺]_i and [Mg²⁺]_i (Fig. 3j, k), implying that the ischaemia-evoked [H⁺]_i rise activates entry of these cations into the cell. Consistent with this, uncaging protons in oligodendrocytes raised [Ca²⁺]_i and [Mg²⁺]_i (Fig. 3l, m).

Few channels allow entry of Ca²⁺ and Mg²⁺ better than Na⁺, but many TRP channels share this property²⁰ (and TRP channel activation can cause ischaemic damage to neurons²¹ and astrocytes²²). Of these channels, only^{20,23,24} TRPA1 and TRPV3 are known to be activated by intracellular H⁺, so we applied modulators of these channels and examined the effect on oligodendrocyte [Ca²⁺]_i (TRP agonist and blocker specificity are discussed in Supplementary Information). The TRPA1/TRPV3 blocker isopentenyl pyrophosphate²⁵ (IPP) and the TRPA1 blocker²⁶ HC-030031 slowed and reduced the [Ca²⁺]_i rise evoked by uncaging H⁺ in the cell (Fig. 3l). The TRPA1/TRPV3 agonists^{20,27} menthol, vanillin, carvacrol (Cv) and 2-APB all raised [Ca²⁺]_i in oligodendrocyte somata and myelinating processes, as did the TRPA1 agonists²⁷ AITC, polygodial and flufenamic acid, while the TRPV3 agonists^{20,27} camphor and farnesyl pyrophosphate (FPP) did not (Fig. 4a, b). Thus, TRPA1 subunit-including channels contribute to these responses, but TRPV3 channels are not needed. The carvacrol-evoked rise of [Ca²⁺]_i was reduced by HC-030031 which blocks TRPA1 but not TRPV3²⁶, by the TRPA1/TRPV3 blocker isopentenyl pyrophosphate²⁵ (IPP), and by TRPA1 knockout (Fig. 4b), again implying involvement of TRPA1 channels. It was unaffected by buffering [H⁺]_i (Fig. 4b), consistent with Ca²⁺ entry via TRPA1 being downstream of the [H⁺]_i rise, as seen with H⁺ uncaging (Fig. 3l).

Using *in situ* hybridization, a TRPA1 probe labelled cerebellar white matter in rat and mouse, while a TRPV3 probe labelled rat only (Extended Data Fig. 5a). Immunocytochemistry revealed that TRPA1- and TRPV3-expressing cells included myelinating oligodendrocytes expressing Olig2 and CC1 (Extended Data Fig. 5b–d).

Consistent with ischaemia raising [Ca²⁺]_i by activating TRPA1-rather than TRPV3-containing channels, the general TRP blockers^{20,27} ruthenium red (RuR, 10 μM) and La³⁺ (1 mM) and the TRPA1/TRPV3 blocker²⁵ IPP (200 μM) reduced the [Ca²⁺]_i rise, as did HC-030031 and A967079, which block TRPA1 but not TRPV3^{26,28} (Fig. 4c–e). Knockout of TRPA1 slowed and halved the ischaemia-evoked [Ca²⁺]_i rise, and the TRPA1/TRPV3 blocker IPP produced no further reduction in the knockout (suggesting no contribution of TRPV3; see Fig. 4f), while knockout of TRPV3 did not affect the [Ca²⁺]_i rise, and the TRPA1 blocker HC-030031 slowed and halved the rise occurring in the TRPV3 knockout (Fig. 4g). Blockers of many other TRP channels had no effect (Extended Data Fig. 6). Thus TRPA1 is the dominant contributor to the ischaemia-evoked rise of [Ca²⁺]_i in oligodendrocytes.

The larger (70%) block of the ischaemia-evoked [Ca²⁺]_i rise by the TRPA1 blocker HC-030031 than by TRPA1 knockout (50%; Figs 4e, f) suggests that there may be compensatory upregulation of another Ca²⁺ entry pathway in the TRPA1 knockout, which normally generates only ~30% of the [Ca²⁺]_i rise. Introducing high pH-buffering-power solution into the cell blocked the [Ca²⁺]_i rise in the TRPA1 knockout (Fig. 4h), implying that the non-TRPA1 Ca²⁺ entry pathway is also H⁺-activated. Since the non-specific TRP blockers RuR and La³⁺ abolished the ischaemia-evoked [Ca²⁺]_i rise, these data suggest that there is another Ca²⁺-permeable TRP channel (neither TRPA1 nor TRPV3) that is activated by internal H⁺ in oligodendrocytes and generates ~30% of the ischaemia-evoked [Ca²⁺]_i rise.

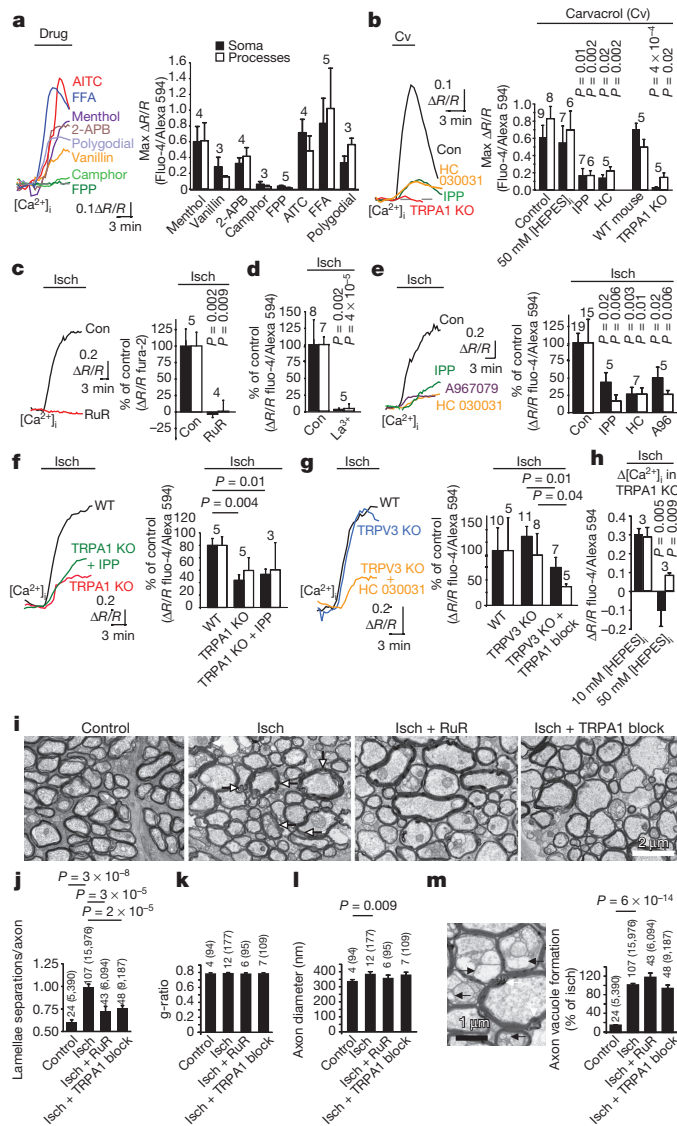


Figure 4 | TRPA1 mediates ischaemic Ca^{2+} accumulation and myelin damage. **a**, $\Delta[Ca^{2+}]_i$ to TRPA1/TRPV3 agonists (mM) (menthol 2, vanillin 1, 2-APB 2), TRPV3 agonists (camphor 2, FPP 0.5) and TRPA1 agonists (AITC 0.5, FFA 1, polygodial 0.2). **b**, $\Delta[Ca^{2+}]_i$ to TRPA1/TRPV3 agonist carvacrol (2 mM, in 1 μ M TTX) is inhibited by TRPA1/TRPV3 antagonist isopentyl pyrophosphate (IPP, 200 μ M, Mann–Whitney test on soma), TRPA1 antagonist HC-030031 (80 μ M), and TRPA1 knock-out (KO) (Mann–Whitney test on processes), but not by 50 mM internal [HEPES] (P values compare with control). **c**, **d**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ is blocked by ruthenium red (RuR, 10 μ M) (**c**), and La^{3+} (**d**, 1 mM, using HEPES-buffered external, Mann–Whitney test on soma). **e**, Block of ischaemia-evoked $\Delta[Ca^{2+}]_i$ by TRPA1/TRPV3 blocker IPP (200 μ M) and TRPA1 blockers HC-030031 (80 μ M) and A967079 (10 μ M). Mann–Whitney P values compare with control. **f**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ in wild-type mice, with TRPA1 knocked out ($P=0.07$ for processes) and with TRPA1/TRPV3 blocker IPP (200 μ M) also present. **g**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ in wild-type mice, with TRPV3 knocked out and with TRPA1 blockers HC-030031 (80 μ M, 6 cells) or A967079 (10 μ M, 1 cell) also present (Mann–Whitney P values). **h**, Ischaemia-evoked $\Delta[Ca^{2+}]_i$ in TRPA1 KO with 10 and 50 mM internal [HEPES]. **i**, Electron microscopy showing control optic nerves and myelin decompaction (white arrows) after 60 min ischaemia or ischaemia in RuR (10 μ M), or A967079 (10 μ M) and HC-030031 (80 μ M) together (TRPA1 block). **j–m**, Lamella separations (**j**), g-ratio (**k**), axon diameter (**l**) and axon vacuoles (**m**, electron microscopy shows vacuoles (black arrows) within axon and periaxonal space) in control, ischaemia alone or ischaemia with RuR (10 μ M) or with A967079 (10 μ M) and HC-030031 (80 μ M) (TRPA1 block). Bar numbers are ‘images (axons)’. P values for **j–m** from Mann–Whitney tests, except **l** from Kolmogorov–Smirnov test. Error bars, s.e.m.; data from rat unless stated otherwise.

To assess the role of TRPA1-containing channels in evoking myelin damage, we exposed rat optic nerves to 60 min ischaemia. This led to disruption of myelin sheaths^{2,3}, which we quantified by counting the regions of myelin decompaction (lamellar separation) per axon cross section (see Methods). Taking as baseline the level of decompaction that occurs in control nerves during processing for electron microscopy, ischaemia increased decompaction ($P=3 \times 10^{-8}$), and ruthenium red or the TRPA1 blockers HC-030031 and A967079 (applied together) reduced this increase by 69% ($P=3 \times 10^{-5}$) and 59% ($P=2 \times 10^{-5}$), respectively (Fig. 4i, j). Ischaemia did not affect the axon g ratio (Fig. 4k; see Methods), but increased axon diameter through swelling (Fig. 4l). It also caused some axon vacuolization (Fig. 4m): vacuoles were seen in 4.3% of 2,390 control axons, but in 21% of 15,976 axons after ischaemia. Vacuolization was not prevented by TRP channel block (Fig. 4m), suggesting different mechanisms for axon and myelin damage.

Thus, ischaemic damage to oligodendrocytes differs fundamentally from that in neurons (Extended Data Fig. 7), where $[Ca^{2+}]_i$ is raised by glutamate-gated receptors (and later by TRP channels activated by reactive oxygen species²¹). Contradicting current ideas^{2–4}, ischaemia does not damage oligodendrocytes by activating Ca^{2+} entry through ionotropic glutamate receptors in their membranes. Instead, ischaemia-evoked sodium pump inhibition and glutamate release evoke a long-lasting rise of $[K^+]_o$ that, together with metabolic changes, acidifies the oligodendrocyte, activating H^+ -gated TRP channels through which Ca^{2+} enters.

In the optic nerve, ischaemia-evoked Ca^{2+} entry into oligodendrocytes is blocked by NMDA receptor antagonists², contradicting our demonstrations that NMDA evokes no $[Ca^{2+}]_i$ rise in oligodendrocytes (Fig. 2) and that the ischaemia-evoked $[Ca^{2+}]_i$ rise is unaffected by NMDA receptor blockers (Fig. 3). Conceivably, in the optic nerve, NMDA receptors on astrocytes²⁹ make a greater contribution than in cerebellum to generating the ischaemia-evoked rise of $[K^+]_o$ and thus the $[Ca^{2+}]_i$ rise.

TRPA1 generates $\sim 70\%$ of the ischaemia-evoked $[Ca^{2+}]_i$ rise, and TRPA1 blockers reduce ischaemic damage to myelin (Fig. 4). Consequently, blocking oligodendrocyte TRPA1-containing channels may reduce myelin loss during the energy deprivation that follows stroke, secondary ischaemia caused by spinal cord injury, or hypoxia in multiple sclerosis³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 September; accepted 7 December 2015.

Published online 13 January 2016.

- Stys, P. K., Ransom, B. R., Waxman, S. G. & Davis, P. K. Role of extracellular calcium in anoxic injury of mammalian central white matter. *Proc. Natl Acad. Sci. USA* **87**, 4212–4216 (1990).
- Micu, I. et al. NMDA receptors mediate calcium accumulation in myelin during chemical ischaemia. *Nature* **439**, 988–992 (2006).
- Salter, M. G. & Fern, R. NMDA receptors are expressed in developing oligodendrocyte processes and mediate injury. *Nature* **438**, 1167–1171 (2005).
- Kárádóttir, R., Cavellier, P., Bergersen, L. H. & Attwell, D. NMDA receptors are expressed in oligodendrocytes and activated in ischaemia. *Nature* **438**, 1162–1166 (2005).
- Bakiri, Y., Hamilton, N. B., Kárádóttir, R. & Attwell, D. Testing NMDA receptor block as a therapeutic strategy for reducing ischaemic damage to CNS white matter. *Glia* **56**, 233–240 (2008).
- McCarran, W. J. & Goldberg, M. P. White matter axon vulnerability to AMPA/kainate receptor-mediated ischemic injury is developmentally regulated. *J. Neurosci.* **27**, 4220–4229 (2007).
- Agrawal, S. K. & Fehlings, M. G. Role of NMDA and non-NMDA ionotropic glutamate receptors in traumatic spinal cord axonal injury. *J. Neurosci.* **17**, 1055–1063 (1997).
- Sarrafzadeh, A. S., Kiening, K. L., Callens, T. A. & Unterberg, A. W. Metabolic changes during impending and manifest cerebral hypoxia in traumatic brain injury. *Br. J. Neurosurg.* **17**, 340–346 (2003).
- Li, S., Mealing, G. A., Morley, P. & Stys, P. K. Novel injury mechanism in anoxia and trauma of spinal cord white matter: glutamate release via reverse Na^+ -dependent glutamate transport. *J. Neurosci.* **19**, RC16 (1999).

10. Back, S. A. *et al.* Hypoxia-ischemia preferentially triggers glutamate depletion from oligodendroglia and axons in perinatal cerebral white matter. *J. Cereb. Blood Flow Metab.* **27**, 334–347 (2007).
11. Yuan, X., Eisen, A. M., McBain, C. J. & Gallo, V. A role for glutamate and its receptors in the regulation of oligodendrocyte development in cerebellar tissue slices. *Development* **125**, 2901–2914 (1998).
12. Lundgaard, I. *et al.* Neuregulin and BDNF induce a switch to NMDA receptor-dependent myelination by oligodendrocytes. *PLoS Biol.* **11**, e1001743 (2013).
13. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
14. De Biase, L. M., Nishiyama, A. & Bergles, D. E. Excitability and synaptic communication within the oligodendrocyte lineage. *J. Neurosci.* **30**, 3600–3611 (2010).
15. Kukley, M., Nishiyama, A. & Dietrich, D. The fate of synaptic input to NG2 glial cells: neurons specifically downregulate transmitter release onto differentiating oligodendroglial cells. *J. Neurosci.* **30**, 8320–8331 (2010).
16. Rossi, D. J., Oshima, T. & Attwell, D. Glutamate release in severe brain ischaemia is mainly by reversed uptake. *Nature* **403**, 316–321 (2000).
17. Limbrick, D. D. Jr, Sombati, S. & DeLorenzo, R. J. Calcium influx constitutes the ionic basis for the maintenance of glutamate-induced extended neuronal depolarization associated with hippocampal neuronal death. *Cell Calcium* **33**, 69–81 (2003).
18. Hamann, M., Rossi, D. J., Mohr, C., Andrade, A. L. & Attwell, D. The electrical response of cerebellar Purkinje neurons to simulated ischaemia. *Brain* **128**, 2408–2420 (2005).
19. Lesage, F. *et al.* TWIK-1, a ubiquitous human weakly inward rectifying K⁺ channel with a novel structure. *EMBO J.* **15**, 1004–1011 (1996).
20. Nilius, B. & Szallasi, A. Transient receptor potential channels as drug targets: from the science of basic research to the art of medicine. *Pharmacol. Rev.* **66**, 676–814 (2014).
21. Aarts, M. *et al.* A key role for TRPM7 channels in anoxic neuronal death. *Cell* **115**, 863–877 (2003).
22. Butenko, O. *et al.* The increased activity of TRPV4 channel in the astrocytes of the adult rat hippocampus after cerebral hypoxia/ischemia. *PLoS One* **7**, e39959 (2012).
23. Wang, Y. Y., Chang, R. B. & Liman, E. R. TRPA1 is a component of the nociceptive response to CO₂. *J. Neurosci.* **30**, 12958–12963 (2010).
24. Cao, X., Yang, F., Zheng, J. & Wang, K. Intracellular proton-mediated activation of TRPV3 channels accounts for the exfoliation effect of α -hydroxyl acids on keratinocytes. *J. Biol. Chem.* **287**, 25905–25916 (2012).
25. Bang, S., Yoo, S., Yang, T. J., Cho, H. & Hwang, S. W. Isopentenyl pyrophosphate is a novel antinociceptive substance that inhibits TRPV3 and TRPA1 ion channels. *Pain* **152**, 1156–1164 (2011).
26. McNamara, C. R. *et al.* TRPA1 mediates formalin-induced pain. *Proc. Natl Acad. Sci. USA* **104**, 13525–13530 (2007).
27. Clapham, D. E. SnapShot: mammalian TRP channels. *Cell* **129**, 220.e1–220.e2 (2007).
28. Chen, J. *et al.* Selective blockade of TRPA1 channel attenuates pathological pain without altering noxious cold sensation or body temperature regulation. *Pain* **152**, 1165–1172 (2011).
29. Hamilton, N. *et al.* Mechanisms of ATP- and glutamate-mediated calcium signaling in white matter astrocytes. *Glia* **56**, 734–749 (2008).
30. Davies, A. L. *et al.* Neurological deficits caused by tissue hypoxia in neuroinflammatory disease. *Ann. Neurol.* **74**, 815–825 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Supported by a Wellcome Trust Programme Grant and Senior Investigator Award, an ERC Advanced Investigator Award, and the EU (Leukotreat). We thank P. Marcaggi for help with ion-sensitive electrodes, and J. Sexton and J. Wood for knockout mice. K. Kolodziejczyk was in the 4 year PhD Programme in Neuroscience at UCL. We thank L. Arancibia-Carcamo, M. Ford, A. Gibb, R. Jolivet, J. Kittler, M. Sajic, A. Silver and K. Smith for comments on the manuscript.

Author Contributions N.B.H. carried out the experiments in Figs 1g, h, 2, 3 and 4, and Extended Data Figs 2, 3, 4 and 6, and contributed to Figs 1a–d, f and Extended Data Fig. 1. K.K. carried out experiments in Fig. 1a–f and Extended Data Fig. 1. E.K. carried out experiments for Extended Data Fig. 5. All authors designed experiments and analysed data. D.A. and N.B.H. wrote the manuscript. All authors revised the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.B.H. (nicola.hamilton-whitaker@ucl.ac.uk) or D.A. (d.attwell@ucl.ac.uk).

METHODS

Animals. Experiments used Sprague-Dawley rats or transgenic mice of either sex. Data are from rats unless stated otherwise. Animal procedures were carried out in accordance with the guidelines of the UK Animals (Scientific Procedures) Act 1986 and subsequent amendments. TRPV3 knockout (KO) mice were obtained from JAX (<http://jaxmice.jax.org/strain/010773.html>). TRPA1 KO mice were obtained as a double knockout with TRPV1 knocked out (kindly provided by J. Wood and J. Sexton). TRPV1 does not contribute to the ischaemia-evoked $[Ca^{2+}]_i$ rise described here because the TRPV1 antagonist^{20,27} capsazepine did not reduce the ischaemia-evoked $[Ca^{2+}]_i$ rise in rat oligodendrocytes (Extended Data Fig. 6a) and the TRPV1 agonists^{20,27} capsaicin (10 μ M) and camphor (2 mM) did not evoke a $[Ca^{2+}]_i$ rise (see Specificity of drugs acting on TRP channels section and Fig. 4a). Wild-type and (double) KO mice were from a colony obtained by breeding mice doubly heterozygous for the TRPA1 and TRPV1 knockouts. The wild-type and KO mice compared shared the same doubly heterozygous grandparents.

Brain slice preparation. Cerebellar slices (225 μ m thick) were prepared from the cerebellum of P12 rats in ice-cold solution containing (mM) 124 NaCl, 26 NaHCO₃, 1 NaH₂PO₄, 2.5 KCl, 2 MgCl₂, 2–2.5 CaCl₂, 10 glucose, bubbled with 95% O₂/5% CO₂, pH 7.4, as well as 1 mM Na-kynurenate to block glutamate receptors. Slices were then incubated at room temperature (21–24 °C) in the same solution until used in experiments. Cerebellar slices from P10–17 mice were prepared in ice-cold solution containing (mM) 87 NaCl, 25 NaHCO₃, 1.25 NaH₂PO₄, 2.5 KCl, 7 MgCl₂, 0.5 CaCl₂, 25 glucose, 75 sucrose, 1 Na-kynurenate and then transferred to the same solution at 27 °C and allowed to cool naturally to room temperature. Only 1 cell was recorded from in each slice.

Cell identification and electrophysiology. Oligodendrocytes, cerebellar granule cells and hippocampal pyramidal cells were identified by their location and morphology. All cells were whole-cell clamped with pipettes with a series resistance of 8–30 M Ω . Electrode junction potentials were compensated. *I*–*V* relations were from responses to 200 ms voltage steps. Unless otherwise indicated, cells were voltage-clamped at –74 mV.

External solutions. Slices were superfused with either bicarbonate-buffered solution containing (mM) 124 NaCl, 2.5 KCl, 26 NaHCO₃, 1 NaH₂PO₄, 2–2.5 CaCl₂, 1 MgCl₂, 10 glucose, pH 7.4, bubbled with 95% O₂ and 5% CO₂, or with HEPES-buffered solution containing (mM) 144 NaCl, 2.5 KCl, 10 HEPES, 1 NaH₂PO₄, 2–2.5 CaCl₂, 1 MgCl₂, 10 glucose, pH set to 7.3 with NaOH, bubbled with 100% O₂. During experiments when NMDA was applied and ion concentration changes were observed with ion-sensitive dyes, MgCl₂ was omitted from the solution to minimize the Mg²⁺ block. For experiments involving Gd³⁺ and La³⁺, the HEPES-based solution was used and NaH₂PO₄ was omitted. To simulate ischaemia we replaced external O₂ with N₂, and external glucose with 7 mM sucrose, added 2 mM iodoacetate to block glycolysis, and 25 μ M antimycin to block oxidative phosphorylation^{4,31}. All ischaemia experiments were done at 33–36 °C, while applications of NMDA and of TRP channel agonists were at 24 °C. Control and drug conditions were interleaved where appropriate. For calcium imaging experiments when applying ischaemia solution to brain slices from transgenic mice, the experimenter was blind to the genotype.

Intracellular solutions. Cells were whole-cell clamped with electrodes containing either Cs- (to improve voltage uniformity) or K-gluconate-based solution, comprising (mM) 130 Cs-gluconate (or K-gluconate), 2 NaCl, 0.5 CaCl₂, 10 HEPES, 10 BAPTA, 2 NaATP, 0.5 Na₂GTP, 2 MgCl₂, 0.5 K-Lucifer yellow, pH set to 7.2 with CsOH or KOH (all from Sigma). The K⁺-based solution was used for current-clamp experiments. For Ca²⁺ imaging experiments, BAPTA was decreased to 0.01 mM and replaced with 10 mM phosphocreatine, added CaCl₂ was reduced to 10 μ M, and Lucifer yellow was replaced with 1 mM Fura-2, or 200 μ M Fluo-4 with 50 μ M Alexa Fluor 594, or 200 μ M X-Rhod-1 with 50 μ M Alexa Fluor 488 (all from Molecular Probes) to allow ratiometric imaging. For imaging pH, Lucifer yellow was replaced with BCECF (96 μ M) and the HEPES concentration was decreased to 0.5 mM. This [HEPES] was also used for control experiments when examining the effect of 50 mM internal [HEPES] on the ischaemia-evoked current; ischaemia-evoked membrane current changes were indistinguishable when 0.5 and 10 mM HEPES were used (see main text), presumably because endogenous pH buffering dominates at these low [HEPES] levels. For experiments where the pH-buffering capacity of the internal solution was increased, 68 mM K-gluconate and 50 mM HEPES were used. When uncaging protons, 2 mM 1-(2-nitrophenyl) ethyl sulphate sodium salt (NPE-caged protons, Tocris) was added to the pipette solution and 10 mM HEPES was replaced by 30 mM Tris to prevent ultraviolet-light-mediated oxidation³² of HEPES (and K-gluconate was reduced from 130 to 120 mM). For Na⁺ and K⁺ imaging experiments, Lucifer yellow was replaced with 1 mM of the Na⁺-sensing dye SBFI tetra-ammonium salt or of the K⁺-sensing dye PBFI tetra-ammonium salt (Molecular Probes). In some experiments MK-801 (1 mM) was added to the internal solution to block NMDA receptors, and cells

were depolarized to –10 mV for 10 s intermittently over a 20 min waiting period to facilitate MK-801 block of open channels.

Single-cell ion imaging and H⁺ uncaging. For Fura-2, SBFI and PBFI imaging when applying NMDA or during ischaemia, white matter oligodendrocytes and grey matter granule cells were patch-clamped with pipettes containing a solution as described above, fluorescence was excited sequentially at 340 \pm 10 nm and 380 \pm 10 nm, and emitted light was collected at 510 \pm 20 nm. The ratio (*R*) of the emission intensities (340 nm/380 nm), after subtraction of the background intensity averaged over 4 distant areas of the image, was used as a measure of intracellular ion concentration. Increases of ion concentration generated a fall of fluorescence (*F*) excited at 380 nm and a rise in fluorescence excited at 340 nm, which is plotted as $\Delta R/R$ in the graphs shown, with $R = F_{340\text{ nm}}/F_{380\text{ nm}}$; an upward deflection corresponds to a rise of concentration of the sensed ion. For Fura-2, SBFI and PBFI, mean values of *R* before applying NMDA or ischaemia solution were 0.41 \pm 0.05 (*n* = 5), 1.68 \pm 0.14 (*n* = 13) and 1.86 \pm 1.10 (*n* = 8) respectively.

Fluo-4 and Alexa Fluor 594 were used in the internal solution to measure $[Ca^{2+}]_i$ changes ratiometrically during H⁺-uncaging and most ischaemia experiments. To measure $[Mg^{2+}]_i$, Mag-Fluo-4 was used instead of Fluo-4. Fluo-4 (or Mag-Fluo-4) and Alexa Fluor 594 fluorescence were excited sequentially every 2, 10 or 30 s at 488 \pm 10 nm and 585 \pm 10 nm, and emission was collected using a triband filter cube (DAPI/FITC/Texas red, 69002, Chroma). The mean ratio of intensities ($F_{488\text{ nm}}/F_{585\text{ nm}}$) before applying NMDA or ischaemia was 0.81 \pm 0.09 (*n* = 16) for Fluo-4 and 0.55 \pm 0.04 (*n* = 6) for Mag-Fluo-4. Caged-H⁺ were uncaged using 380 \pm 20 nm light for 1 s every 2 s (repeated 30 times) interspersed with the above excitation wavelengths. BCECF was imaged every 30 s at 400 and 480 nm, with emission collected using the above tri-band filter. The ratio (*R*) of the emitted light excited by these two wavelengths ($F_{480\text{ nm}}/F_{400\text{ nm}}$) was used as a measure of $[H^+]_i$ (mean value before ischaemia was 17.6 \pm 0.9, *n* = 8) but, since this ratio decreases with increasing $[H^+]_i$, when plotting changes in $\Delta R/R$ in Fig. 3e and Extended Data Fig. 4 we multiplied them by –1 to produce a trace that increased with $[H^+]_i$.

During ischaemia, slices swelled at the time of the anoxic depolarization. When cells were patch-clamped with calcium dyes, the resulting movement of the cell away from the electrode sometimes caused $[Ca^{2+}]_i$ oscillations within the cells. These oscillations did not occur if the patch pipette was removed (after 2 min to allow dye-filling) before the ischaemic solution was applied. Without the pipette attached to the cell, the time-course of the ischaemia-evoked $[Ca^{2+}]_i$ rise was the same as with the electrode attached, but its amplitude was 69% larger (ratio increase 0.21 \pm 0.03, *n* = 20 versus 0.12 \pm 0.02, *n* = 16). In some experiments (those in Fig. 4c–g and Extended Data Fig. 6) we therefore removed the pipette for calcium-imaging.

Control experiments were carried out to check whether the ischaemia-evoked change of pH would affect our $[Ca^{2+}]_i$ measurements. The internal solution for Ca²⁺-sensing was studied in the experimental bath that the slices usually are placed in. The resting ratio of Fluo-4 fluorescence to Alexa 594 fluorescence was not significantly affected by altering the pH of the solution from 7.05 to 6.55, and this also did not affect the change of ratio produced by adding 200 nM Ca²⁺ to the sensing solution (Extended Data Fig. 4b, c). Thus, even a 0.5 unit pH change occurring in the oligodendrocyte would not significantly affect the calcium dye measurements. **AM dye loading.** X-Rhod-1-AM (38 μ M) dye loading with the myelin marker DIOC6 into P12 cerebellar slices was performed as described previously for optic nerves². Loading times ranged from 1–2 h and a de-esterification period of 30 min at 36 °C was allowed before imaging.

Potassium electrodes. Potassium electrodes were made as described³³. Electrodes were pulled with a resistance of 4–10 M Ω . Electrode tips were silanized by heating them to 250 °C for 7 min while N₂ and *N,N*-dimethyltrimethylsilylamine (Fluka) were gassed into the tip from the back of the electrode. The tip was then filled with either 6% valinomycin, 1.5% potassium tetrakis(4-chlorophenyl)borate (Fluka) and 92.5% 1,2-dimethyl-3-nitrobenzene (Fluka) or the pre-made potassium sensitive ionophore I – cocktail B (Fluka). The electrodes were back-filled with the bicarbonate-buffered external solution mentioned above (2.5 mM K⁺), and attached to a sensitive high resistance electrometer (Model FD 223, World Precision Instruments). A reference electrode tip was placed less than 5 μ m away from the K⁺ electrode tip, and the voltage changes measured by it were subtracted from those measured with the K⁺ electrode. $[K^+]_o$ was determined by calibrating each electrode at the end of every experiment with at least 3 different K⁺ concentrations (1, 2.5, 5, 7.5, 10 or 17.5 mM). To check for cross-reactivity, the $[NaCl]_o$ was decreased by 60 mM which led to a –2.2 \pm 0.1 mV change in voltage (*n* = 3), while a pH change from 7.3 to 6.5 led to a 0.47 \pm 0.22 mV change (*n* = 3). Both of these changes are much less than the 17.5 \pm 0.5 mV change (*n* = 18) seen in response to an increase of $[K^+]_o$ from 2.5 to 5 mM (which is consistent with the electrodes used having an average calibration slope of 60.9 \pm 0.9 mV (*n* = 6) per tenfold change of $[K^+]_o$).

Drugs used. Stock solutions of the following drugs were made up in water: NMDA, AP5, NBQX, MK801, 7-CK, TTX, PDC, IPP, CPG, SKF 96365 and RuR. (S)-MCPG and amiloride were made up in external solution. Carvacrol was made up in ethanol. Bicuculline, bumetanide, HC 030031, A967079, flufenamic acid, capsazepine, FTY720-HCl, 2-APB, AITC, RN1734 and ML204 were made up in DMSO. When used, DMSO and ethanol were also added to control solution at the same concentrations, and did not evoke $[Ca^{2+}]_i$ changes at the concentrations used. Stocks were kept at $-20^{\circ}C$ apart from carvacrol, menthol, vanillin, AITC, and RuR, which were made up fresh on each day of use. To minimise evaporation of carvacrol, vanillin and menthol, lids were kept on until the solutions were used. Gd^{3+} and La^{3+} were applied (as chloride salts) in bicarbonate- and phosphate-free solution to avoid chelation by these anions (see External solutions section earlier).

Immunohistochemical labelling of oligodendrocytes. Cerebellar slices were fixed for 30 min in 4% paraformaldehyde (PFA), and incubated for 1 to 6 h in 0.1% Triton X-100, 10% goat serum in phosphate-buffered saline at $21^{\circ}C$, then with primary antibody at $4^{\circ}C$ overnight with agitation, and then 2 h or overnight at $24^{\circ}C$ with secondary antibody. Primary antibodies were: anti-CC1 (mouse, 1:300, Calbiochem OP80 monoclonal) and anti-Olig-2 (rabbit, 1:700, Millipore, AB9610 polyclonal). Secondary antibodies were: goat anti-rabbit Alexa Fluor 488 or 568 (Molecular Probes, 1:1,000), donkey anti-rabbit Alexa Fluor 488 (Millipore, 1:1,000), and goat anti-mouse Alexa Fluor 568 (Millipore, 1:1,000).

Antibody labelling and *in situ* hybridization for TRPA1 and TRPV3. TRPA1 and TRPV3 antibodies appeared to label the myelinating processes and somata of oligodendrocytes in rat, but the labelling was not significantly different in wild-type mice and mice with TRPA1 or TRPV3 knocked out (data not shown). We therefore turned to *in situ* hybridization.

Solutions used for *in situ* hybridization were pretreated with 0.1% DEPC. Animals were perfused with PBS followed by 4% PFA. Brains were post-fixed in 4% PFA overnight at $4^{\circ}C$, cryoprotected in 20% sucrose overnight at $4^{\circ}C$ and frozen in Tissue-Tek OCT Sections (20 μ m) collected onto Superfrost Plus microscope slides (VWR International) were hybridized at $65^{\circ}C$ overnight with hybridization buffer [50% v/v deionized formamide (Sigma), 10% w/v dextran sulphate (Fluka), 0.1 mg ml $^{-1}$ yeast tRNA (Roche), 1 \times Denhardt's solution (Sigma) and 1 \times 'salts' (200 mM NaCl, 5 mM EDTA, 10 mM Tris-HCl pH 7.5, 5 mM NaH $_2$ PO $_4$, 5 mM Na $_2$ HPO $_4$)] containing digoxigenin (DIG)-labelled antisense RNA probe (1:1,000). Sections were washed with a washing solution (50% v/v formamide, 1 \times SSC, 0.1% Tween 20) three times at $65^{\circ}C$ for 30 min, followed by two 1 \times MABT (100 mM maleic acid, 150 mM NaCl, pH 7.5, 0.1% Tween-20) washes at room temperature for 30 min each. Sections were subsequently blocked with blocking solution (2% w/v blocking reagent (Roche Diagnostics), 10% v/v heat-inactivated sheep serum (Sigma) in 1 \times MABT) for 1 h at room temperature and incubated with anti-DIG antibody conjugated with alkaline phosphatase (AP) (Roche Diagnostics, 1:1,500 in blocking solution) at $4^{\circ}C$ overnight. Sections were then washed in 1 \times MABT 5 times for 20 min each at room temperature, followed by two 5 min washes in staining buffer (100 mM NaCl, 50 mM MgCl $_2$, 100 mM Tris-HCl, pH 9.5, 0.1% Tween-20). Development was performed at $37^{\circ}C$ for 24–48 h overnight with nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate in freshly prepared staining solution (50% v/v staining buffer, 25 mM MgCl $_2$, 5% w/v polyvinyl alcohol). Sections were washed in PBS and immunohistochemistry was performed as described above. The plasmids used to generate RNA probes were: IMAGE clone 40129486 for *Trpa1* (linearized with ClaI and transcribed with T3 RNA polymerase) and IMAGE clone 40047664 for *Trpv3* (linearized with XhoI and transcribed with SP6 RNA polymerase). *In situ* hybridization was repeated using at least three animals for each probe.

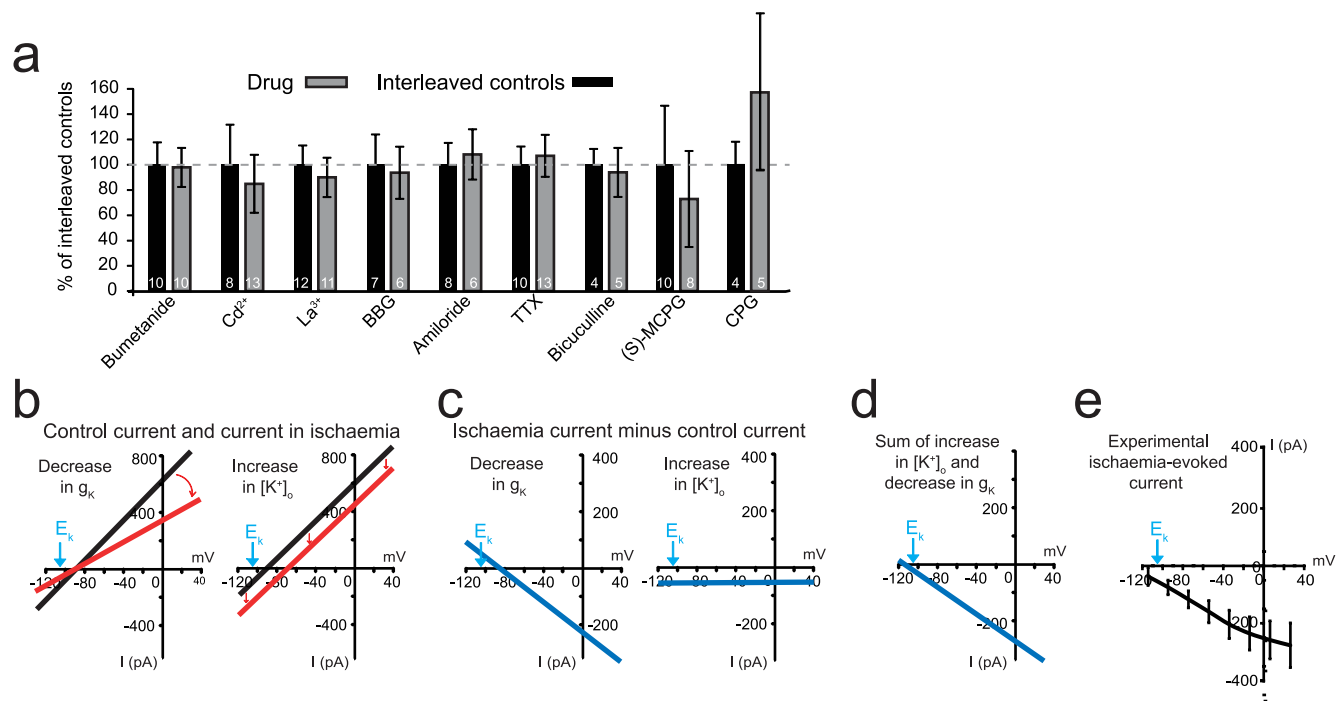
Quantifying myelin decompaction during chemical ischaemia using electron microscopy. For chemical ischaemia experiments, optic nerves were dissected from P28 Sprague-Dawley rats and incubated for 1 h at $36^{\circ}C$ in either control or ischaemic solution with and without the TRPA1/V3 channel blocker ruthenium red (10 μ M) or the combined presence of the TRPA1 blockers HC-030031 (80 μ M) and A967079 (10 μ M). The optic nerves were then immersion fixed in 2% paraformaldehyde and 2% glutaraldehyde in 0.1 M cacodylate buffer overnight.

All samples were then post-fixed in 1% OsO $_4$ /0.1 M cacodylate buffer (pH 7.3) at $3^{\circ}C$ for 2 h before washing in 0.1 M cacodylate buffer (pH 7.3). The samples were dehydrated in a graded ethanol-water series at $3^{\circ}C$ and infiltrated with Agar 100 resin mix. The nerve was then cut transversely at the mid-point, blocked out and hardened. Ultra-thin sections were taken, 300 μ m from the cut end of the middle of the nerve, on a Reichert Ultracut S microtome. Sections were collected on 300 mesh copper grids and stained with lead citrate. The sections were imaged using a Joel 1010 transition electron microscope and a Gatan Orius camera.

In 3 out of 4 experiments the experimenter was blinded to the drug condition before imaging (all 4 experiments gave similar results). One section was used from each nerve and eight 21.5 μ m \times 17.3 μ m images were collected at \times 8,000 magnification, four from the peripheral borders of the nerve at 0° , 90° , 180° and 270° positions on the section, and four covering the central portion of the nerve. In all experiments the image identities were then blinded before analysis, and the number of large separations of lamellae (decompaction) was counted. Decompaction was defined as a visible white inter-lamellar gap being present between at least 2 normal lamellae. Regions of decompaction were normally separated from each other by an area of compact myelin, but when most of the myelin surrounding an axon had separated lamellae, decompacted regions were counted at 0.5 μ m intervals around the sheath. The number of decompacted regions was normalized to the number of axons per image. Some decompaction occurred even in control nerves as a result of the processing for electron microscopy, so we assessed drug block of decompaction by quantifying the ischaemia-evoked increase in decompactions seen without and with the drug present. Myelin g ratios were calculated as the square root of the ratio of the area of the axon to the area of the axon plus myelin sheath. When drawing lines around the axon and sheath, areas of decompaction were ignored, that is, we interpolated the lines from regions that were not decompacted. Axon diameter was calculated as $(4(\text{axon area})/\pi)^{0.5}$. Axon vacuolization was defined as the inclusion of one or more large ($>0.1 \mu$ m) empty membrane bound (often circular) organelles within the axon or periaxonal space (Fig. 4m) which may reflect rearrangement of internal axonal membranes or be formed from inclusion of myelin membranes into the axon (Fig. 4m).

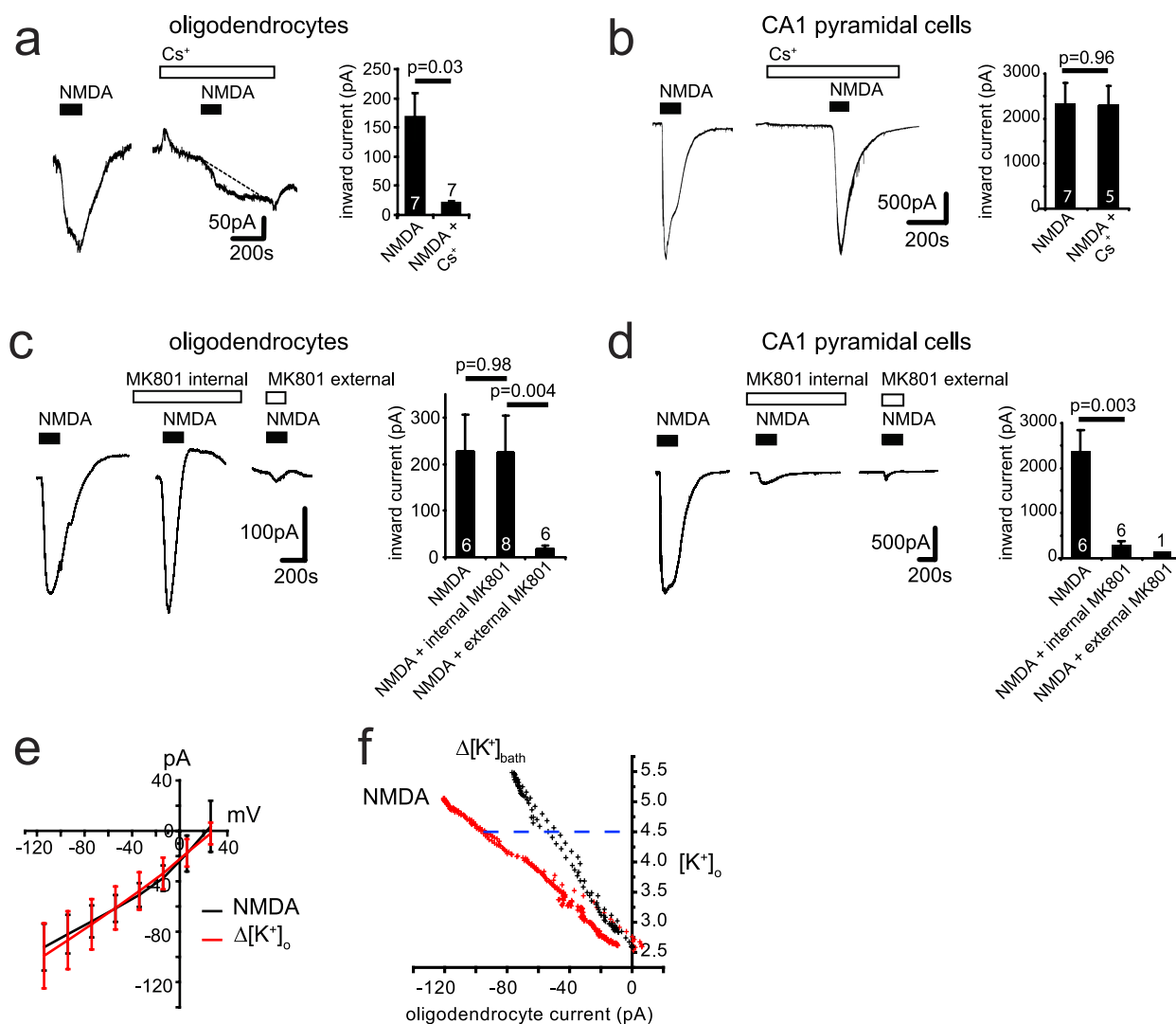
Statistics. Data are presented as mean \pm s.e.m. Experiments were carried out on brain slices from at least 3 animals on at least 3 separate days, except for a few experiments using expensive drugs which were done on only 2 days. Only 1 cell was recorded from in each slice, so the numbers of cells given are also the numbers of slices. *P* values are from two-tailed Student's *t*-tests (for normally distributed data, assessed using Shapiro-Wilk tests) or Mann-Whitney *U* tests (for non-normally distributed data). Normally distributed data were tested for equal variance (*P* < 0.05, unpaired *F*-test) and homo- or heteroscedastic *t*-tests were chosen accordingly. *P* values in the text are from unpaired *t*-tests unless otherwise stated. When small sample sizes (*n* \leq 4) achieved *P* < 0.05, analysis of sample and effect size typically demonstrated a power for detecting the observed effect of 80–99% (mean 92%), with two exceptions: the process data in Fig. 4c (power 78%) and the soma data in Fig. 4h (power 75%). For multiple comparisons within one experiment (usually one figure panel, but measurements of $[Ca^{2+}]_i$ in somata and processes were treated as separate experiments even when plotted in the same figure panel), *P* values were corrected using a procedure equivalent to the Holm-Bonferroni method (for *N* comparisons, the most significant *P* value is multiplied by *N*, the 2nd most significant by *N* – 1, the 3rd most significant by *N* – 2, etc.; corrected *P* values are significant if they are less than 0.05). All statistical analysis was conducted using OriginLab software.

- Allen, N. J., Káradóttir, R. & Attwell, D. A preferential role for glycolysis in preventing the anoxic depolarization of rat hippocampal area CA1 pyramidal cells. *J. Neurosci.* **25**, 848–859 (2005).
- Keynes, R. G., Griffiths, C. & Garthwaite, J. Superoxide-dependent consumption of nitric oxide in biological media may confound *in vitro* experiments. *Biochem. J.* **369**, 399–406 (2003).
- Marcaggi, P., Jeanne, M. & Coles, J. A. Neuron-glia trafficking of NH $_4^+$ and K $^+$: separate routes of uptake into glial cells of bee retina. *Eur. J. Neurosci.* **19**, 966–976 (2004).



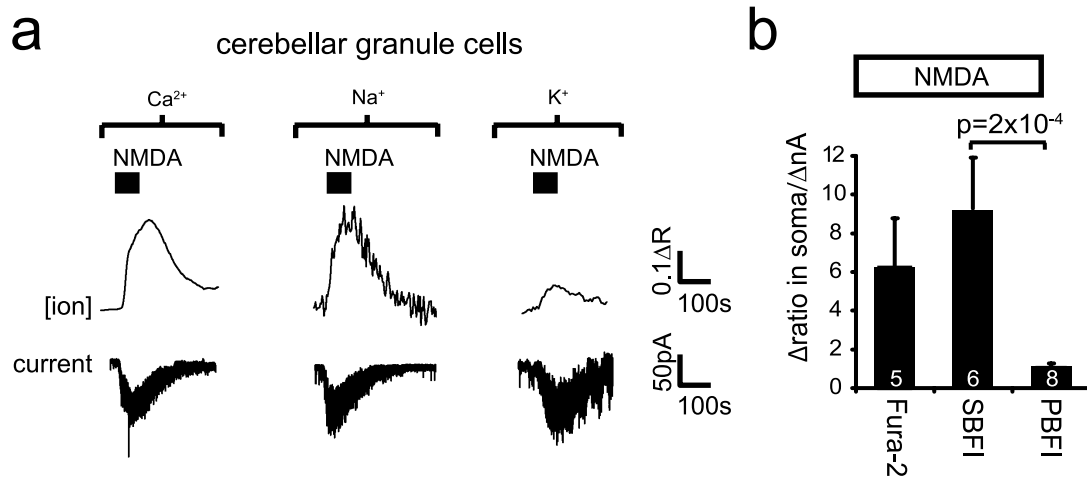
Extended Data Figure 1 | Tests for causes of the ischaemia-evoked current. **a**, Effect of blocking various putative glutamate release mechanisms (blocker concentrations given in Supplementary Information) on peak ischaemia-evoked currents measured in the presence of each drug and in interleaved controls (data from rat). No significant differences were measured ($P > 0.20$). **b**, Schematic showing effect of ischaemia-evoked decrease in resting conductance (which is dominated by g_K , left) and ischaemia-evoked $[K^+]_o$ rise (right) on oligodendrocyte membrane

current. Black lines are control $I-V$ relations. Red lines are $I-V$ relations in ischaemia showing the effect of a conductance decrease (left) or of a positive shift of reversal potential due to $[K^+]_o$ rising (right). **c**, Ischaemia-evoked current change for the two mechanisms in **b** (cf. Fig. 1g). **d**, Sum of currents in **c** gives an $I-V$ relation with a reversal potential more negative than E_K . **e**, Experimentally observed ischaemia-evoked current in 10 oligodendrocytes with 10 mM internal HEPES (difference of curves in Fig. 1f). Error bars are s.e.m.



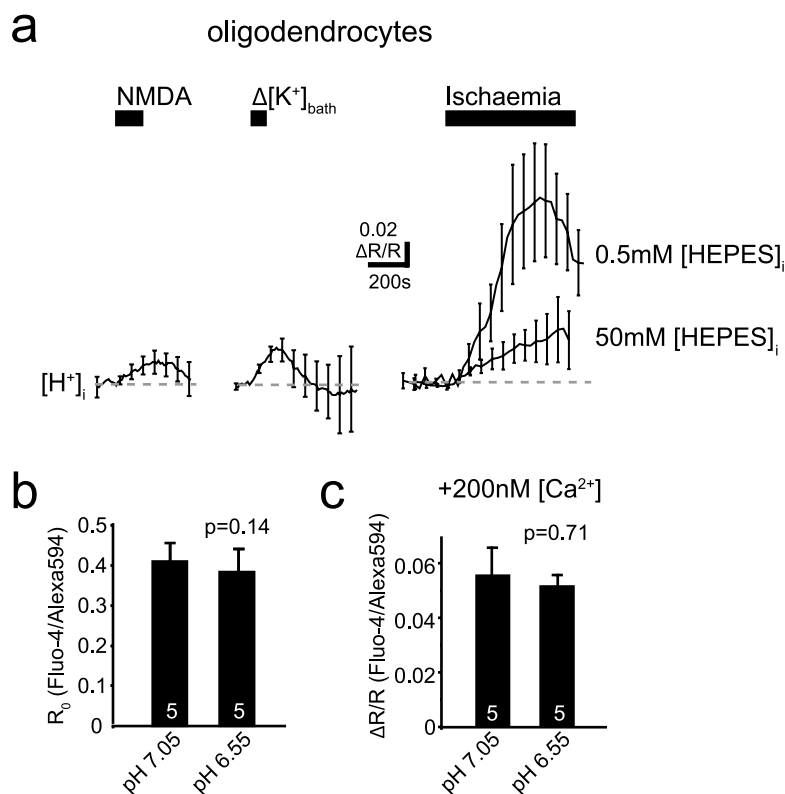
Extended Data Figure 2 | K⁺ flux changes generate the oligodendrocyte NMDA-evoked current. **a, b**, Extracellular Cs⁺ (30 mM, replacing Na⁺) reduces the inward current evoked by 100 μ M NMDA at -74 mV in rat oligodendrocytes (**a**), but not in hippocampal CA1 pyramidal neurons (**b**). **c, d**, Intracellular MK-801 (1 mM) has no effect on NMDA-evoked currents in oligodendrocytes (**c**) but blocks them in pyramidal cells (**d**), while extracellular MK-801 (50 μ M) blocks both. **e**, Voltage-dependence

of the current evoked in 16 oligodendrocytes by 100 μ M NMDA and by elevating $[K^+]_o$ from 2.5 to 5 mM. **f**, Specimen plot of membrane current in an oligodendrocyte versus local $[K^+]_o$ in response to applying 100 μ M NMDA or elevating $[K^+]_o$ from 2.5 to 5 mM. Horizontal cell line shows that if NMDA raises $[K^+]_o$ to (say) 4.5 mM, the current attributable to the $[K^+]_o$ rise alone is 51% of the NMDA-evoked current. Mean value in 11 cells was 49% (see Supplementary Information). Error bars are s.e.m.



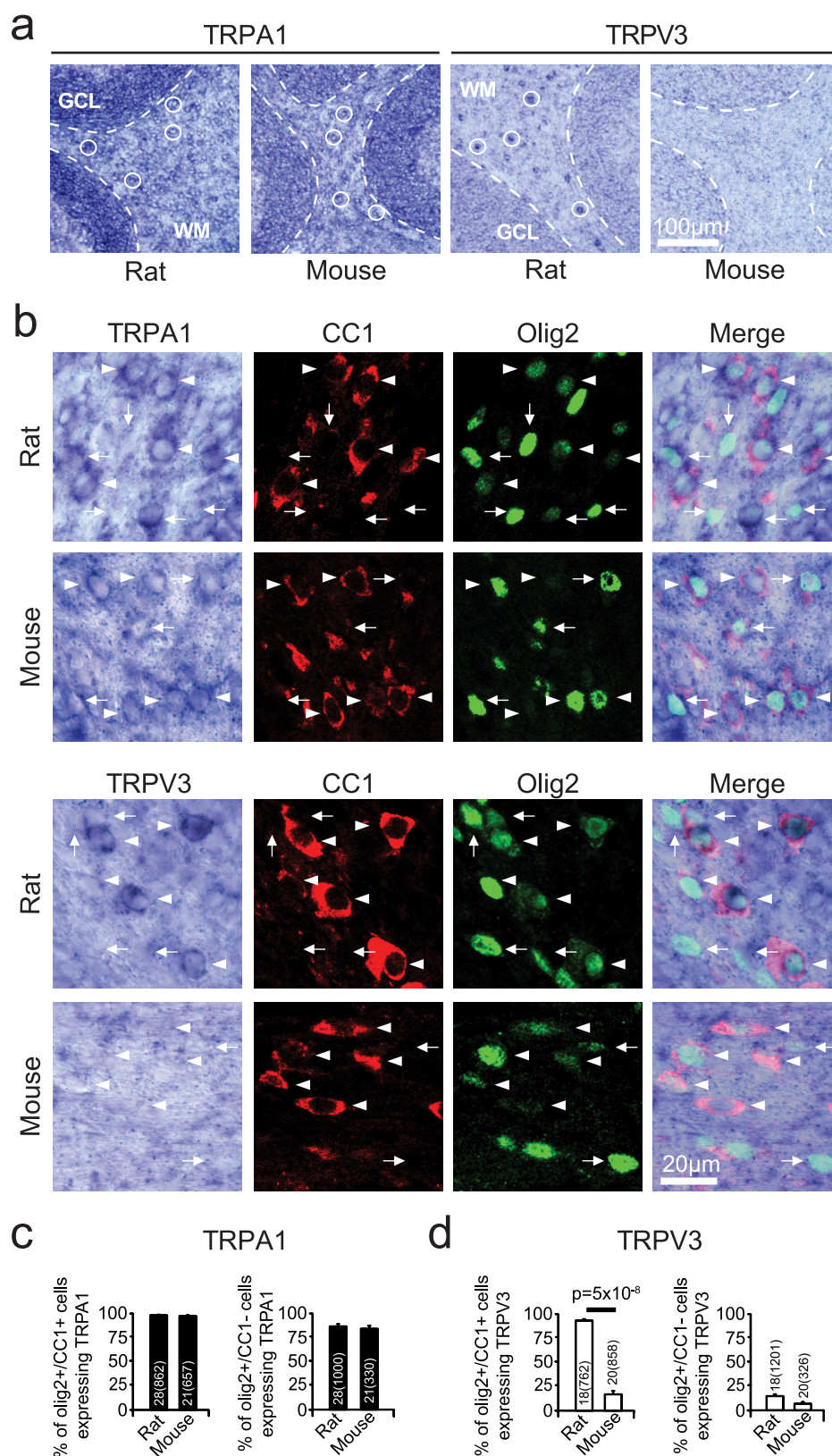
Extended Data Figure 3 | NMDA-evoked ion concentration changes in neurons differ from those in oligodendrocytes. a, Specimen records of rat cerebellar granule cell membrane current and background-subtracted fluorescent dye ratio (R , see Methods) when measuring $[Ca^{2+}]_i$ with Fura-2, $[Na^+]_i$ with SBFI, and $[K^+]_i$ with PBFI, when $100\mu M$ NMDA was

applied. The rise of $[K^+]_i$ seen reflects K^+ entry: $[K^+]_{pipette}$ was 32.5 mM , so $E_K > -60\text{ mV}$ for $[K^+]_o > 3.3\text{ mM}$. **b,** Mean peak fluorescence change normalized to evoked current (number of cells on bars; P value from Mann–Whitney test). Oligodendrocyte data for comparison are shown in Fig. 2. Error bars are s.e.m.



Extended Data Figure 4 | Comparison of NMDA-, $[K^+]_{\text{bath}}$ - and ischaemia-evoked changes of $[H^+]_i$. **a**, Measurements of changes of ratio (R , see Methods) of background-subtracted BCECF fluorescence in rat oligodendrocytes in response to $100\text{ }\mu\text{M}$ NMDA (5 cells) and raising $[K^+]_{\text{bath}}$ from 2.5 to 5 mM with 0.5 mM internal HEPES (6 cells), and

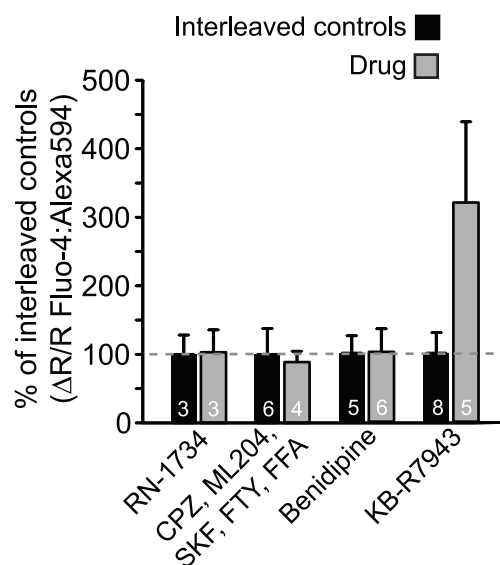
to ischaemia with 0.5 mM and 50 mM internal HEPES (9 and 6 cells, respectively). **b**, **c**, Effect of pH of 10 mM HEPES internal solution on baseline ratio of Fluo-4 to Alexa 594 fluorescence (P value from Mann–Whitney test) (**b**), and change of ratio when $[Ca]$ was increased by 200 nM (**c**). Error bars are s.e.m.



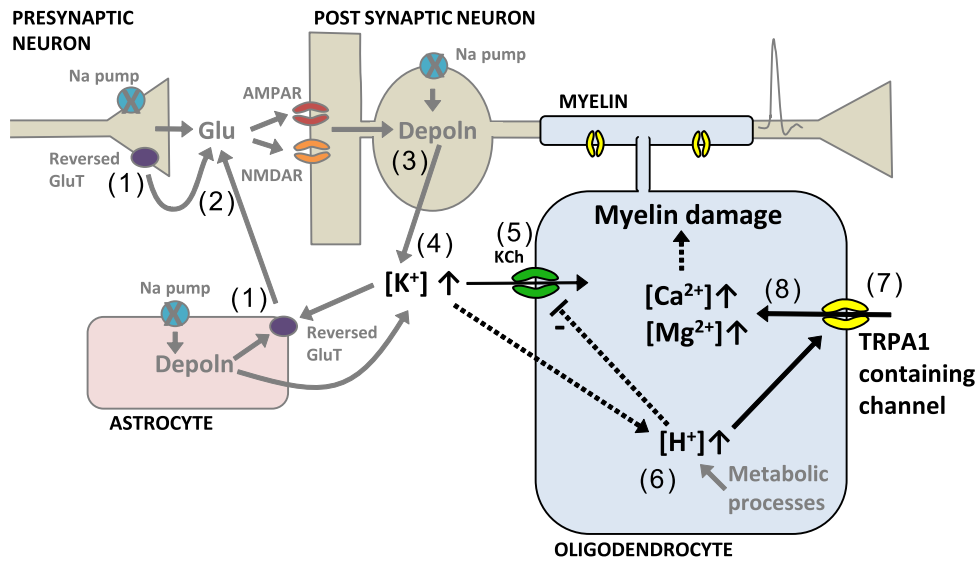
Extended Data Figure 5 | *In situ* hybridization data on TRP channel expression. **a**, *In situ* data for TRPA1 and TRPV3 in the cerebellum of rats and mice show TRPA1 messenger RNA in white matter (WM) cells in rats and mice (with denser expression in the adjacent granule cell layer, GCL), but TRPV3 mRNA only in white matter cells in rats. Specimen cells are labelled with white circles. **b**, Higher magnification views of white matter, combining *in situ* hybridization for TRPA1 and TRPV3 with immunocytochemistry for Olig2 (to label oligodendrocyte lineage

cells) and CC1 (to define myelinating oligodendrocytes). TRPA1 mRNA is present (in rats and mice) and TRPV3 mRNA is present (in rats but not mice) in myelinating oligodendrocytes (Olig2⁺, CC1⁺; arrowheads) and also in some presumed oligodendrocyte precursor cells (Olig2⁺, CC1⁻; arrows). **c**, **d**, Quantification of presence of mRNA for TRPA1 (**c**) and TRPV3 (**d**) in different oligodendrocyte lineage cell classes. Numbers on bars are 'images analysed (cells counted)'. *P* value from Mann–Whitney test. Error bars are s.e.m.

oligodendrocytes



Extended Data Figure 6 | Further evidence for the identity of TRP channels in oligodendrocytes. $[Ca^{2+}]_i$ increase (ratio signal from Fluo-4 and Alexa Fluor 594) in rat oligodendrocyte somata when ischaemia solution was applied with the following drugs present (data normalized to interleaved controls, shown as black bars): RN-1734 (0.5 mM), which blocks TRPV4 and, less well, TRPV1, TRPV3 and TRPM8; a cocktail of blockers inhibiting (see Supplementary Information section on Specificity of drugs acting on TRP channels section) TRPP2, TRPC3, TRPC4, TRPC5, TRPC6, TRPC7, TRPM2, TRPM4, TRPM5, TRPV1, TRPV2, TRPM7, TRPM8 and TRPP1, as well as the store-operated calcium channel component STIM1 and some voltage-gated calcium channels; blocking voltage-gated Ca^{2+} channels with 10 μ M benidipine; or blocking reversed Na/Ca exchange with 10 μ M KB-R7943 mesylate (P values, from Mann-Whitney test and t -test as appropriate, were non-significant ($P > 0.28$)). Error bars are s.e.m.



Extended Data Figure 7 | Schematic of how oligodendrocyte $[Ca^{2+}]_i$ is raised in ischaemia. Run-down of transmembrane ion gradients when Na pump stops (crosses) leads to glutamate transporters (GluT) reversing (1) and releasing glutamate (2). This depolarizes (and causes a neurotoxic $[Ca^{2+}]_i$ rise in) neurons (3) and raises $[K^+]_o$ (4), causing an inward current (5) through oligodendrocyte K⁺ channels (KCh). At the

same time, metabolic changes and also the rise of $[K^+]_o$ lead to a rise in oligodendrocyte $[H^+]_i$ (6). This decreases the membrane K⁺ conductance (either directly or via TRPA1-containing channels opening) which contributes to the inward current generated, and opens TRPA1-containing channels (7) that let Ca²⁺ and Mg²⁺ into the cell (8). The resulting rise of $[Ca^{2+}]_i$ damages the myelin.

Mitofusin 2 maintains haematopoietic stem cells with extensive lymphoid potential

Larry L. Luchsinger^{1,2}, Mariana Justino de Almeida^{1,3}, David J. Corrigan^{1,3}, Melanie Mumau^{1,3} & Hans-Willem Snoeck^{1,2,3}

Haematopoietic stem cells (HSCs), which sustain production of all blood cell lineages¹, rely on glycolysis for ATP production^{2,3}, yet little attention has been paid to the role of mitochondria. Here we show in mice that the short isoform of a critical regulator of HSCs, *Prdm16* (refs 4, 5), induces mitofusin 2 (*Mfn2*), a protein involved in mitochondrial fusion and in tethering of mitochondria to the endoplasmic reticulum. Overexpression and deletion studies, including single-cell transplantation assays, revealed that *Mfn2* is specifically required for the maintenance of HSCs with extensive lymphoid potential, but not, or less so, for the maintenance of myeloid-dominant HSCs. *Mfn2* increased buffering of intracellular Ca^{2+} , an effect mediated through its endoplasmic reticulum-mitochondria tethering activity^{6,7}, thereby negatively regulating nuclear translocation and transcriptional activity of nuclear factor of activated T cells (Nfat). Nfat inhibition rescued the effects of *Mfn2* deletion in HSCs, demonstrating that negative regulation of Nfat is the prime downstream mechanism of *Mfn2* in the maintenance of HSCs with extensive lymphoid potential. Mitochondria therefore have an important role in HSCs. These findings provide a mechanism underlying clonal heterogeneity among HSCs^{8–11} and may lead to the design of approaches to bias HSC differentiation into desired lineages after transplantation.

Within the haematopoietic system, the transcriptional co-regulator, *Prdm16*, is expressed selectively in HSCs and its deletion severely impairs HSC maintenance^{4,5}. However, its molecular targets remain unknown. We observed that in *Prdm16*^{-/-} HSCs (Lin⁻Sca1⁺Kit⁺CD48⁻CD150⁺Flt3⁻, Extended Data Fig. 1) and mouse embryonic fibroblasts (MEFs) mitochondria were fragmented (Fig. 1a and Extended Data Fig. 2a). Mitochondria undergo dynamic fusion and fission^{12,13}. Fusion is driven by the outer membrane GTPases, mitofusin (*Mfn*) 1 and 2, and by the inner membrane protein optic atrophy 1 (*Opa1*), while fission requires dynamin-related protein (*Drp1*)^{12,13}. Culture of *Prdm16*^{-/-} MEFs in the presence of the *Drp1* inhibitor, mDivi1 (ref. 14), restored mitochondrial length (Extended Data Fig. 2b, c), suggesting a fusion defect, which was further documented in a mitochondrial fusion assay (Extended Data Fig. 2d). Expression of *Mfn2* protein and of *Mfn2*, but not *Mfn1*, messenger RNA was lower in *Prdm16*-deficient MEFs and HSCs compared to wild type (Fig. 1b–d and Extended Data Fig. 2e). Furthermore, mitochondria were similarly fragmented in *Prdm16*^{-/-} and in *Mfn2*^{-/-} MEFs (Extended Data Fig. 2f) and mitochondrial length was restored after lentiviral expression of *Mfn2* in *Prdm16*^{-/-} MEFs (Fig. 1e, f and Extended Data Fig. 2g), suggesting that *Mfn2* is a target of *Prdm16*.

Prdm16 exists in two isoforms arising from distinct transcription start sites, full-length (fl) and short (s) *Prdm16*, which lacks the amino-terminal PR-domain (Extended Data Fig. 3a)^{15,16}. Only s*Prdm16*, but not fl*Prdm16*, activated a *Mfn2* promoter luciferase reporter (Extended Data Fig. 3b, c), and induced *Mfn2* mRNA in *Prdm16*^{-/-} MEFs (Fig. 1g). Consistent with these findings, chromatin

immunoprecipitation in MEFs using Flag-tagged isoforms of *Prdm16* showed binding of s*Prdm16*, but not of fl*Prdm16*, to the *Mfn2* promoter (Extended Data Fig. 3d)¹⁷. *Mfn2* is therefore a direct target of s*Prdm16*. Although deletion of *Prdm16* did not affect *Mfn1* (Fig. 1b, c), transduction of s*Prdm16* did increase *Mfn1* mRNA expression (Fig. 1g). *Mfn1* is therefore susceptible to regulation by s*Prdm16*, but with a higher and probably unphysiological threshold. Lentiviral transduction of *Mfn2* did not rescue the competitive repopulation defect of *Prdm16*^{+/-} HSCs, however (Extended Data Fig. 3e), indicating that multiple components of the s*Prdm16* and fl*Prdm16* transcriptional program are required for HSC maintenance.

Induction of *Mfn2* by *Prdm16*, a critical regulator of HSCs, suggests a role for *Mfn2* in HSC function. We therefore assessed mitochondrial length and *Mfn2* expression in haematopoietic cells. HSCs display clonal heterogeneity in their differentiation potential ranging from rare lymphoid-biased HSCs, to balanced myeloid/lymphoid and myeloid-dominant HSCs with low lymphoid potential^{8–11}. Although the underlying mechanism is unknown and neither functional nor phenotypic classifications are absolute, myeloid-dominant HSCs are enriched in the CD150^{hi}, while HSCs with extensive lymphoid potential are enriched in the CD150^{lo} fraction^{18,19}. HSCs expressed more *Mfn2* mRNA (Fig. 2a) and protein (Fig. 2b) than more mature populations. Within the HSC compartment, CD150^{lo} HSCs expressed more *Mfn2* mRNA (Fig. 2a) and protein (Fig. 2c) than did CD150^{hi} HSCs. In contrast, *Mfn1* did not show HSC-selective expression, and its expression in CD150^{lo} HSCs was tenfold lower than that of *Mfn2* (Fig. 2a). In accordance with *Mfn2* induction by s*Prdm16*, s*Prdm16* was the predominant *Prdm16* isoform in CD150^{lo} but not in CD150^{hi} HSCs (Fig. 2d). Using mice expressing a mitochondrially targeted Dendra2 fluorescent protein (Pham mice)²⁰, we observed longer mitochondria in HSCs compared to other haematopoietic populations, and within the HSC compartment, in CD150^{lo} than in CD150^{hi} cells (Fig. 2e and Extended Data Fig. 4a, b). Mitochondrial length therefore paralleled *Mfn2* expression.

As these findings suggested a subpopulation-specific role for *Mfn2* in HSCs, we examined mice with conditional deletion of *Mfn2* (ref. 21) in the haematopoietic system (*Mfn2*^{fl/fl}-Vav-Cre). The frequency of progenitors in the bone marrow and thymus and of mature populations in blood and spleen were similar in *Mfn2*^{fl/fl}-Vav-Cre mice and *Mfn2*^{fl/fl} littermates (Extended Data Table 1). The Lin⁻Sca1⁺Kit⁺CD48⁻CD150⁺ HSC compartment in *Mfn2*^{fl/fl}-Vav-Cre mice showed mitochondrial fragmentation (Extended Data Fig. 5a, b), was smaller (Extended Data Table 1) and expressed more CD150 (Extended Data Fig. 5c) compared to that of *Mfn2*^{fl/fl} mice, indicating a loss primarily of CD150^{lo} HSCs. Competitive repopulation studies²² showed a further increase in CD150 expression within the donor HSC compartment (Extended Data Fig. 5d, e) and a defect in long-term lymphoid repopulation in recipients of *Mfn2*^{fl/fl}-Vav-Cre adult bone marrow (Fig. 3a) and fetal liver cells (Extended Data Fig. 5f). A decrease in myeloid repopulation was noted, but did not

¹Columbia Center for Translational Immunology, Columbia University Medical Center, New York, New York 10032, USA. ²Department of Medicine, Columbia University Medical Center, New York, New York 10032, USA. ³Department of Microbiology and Immunology, Columbia University Medical Center, New York, New York 10032, USA.

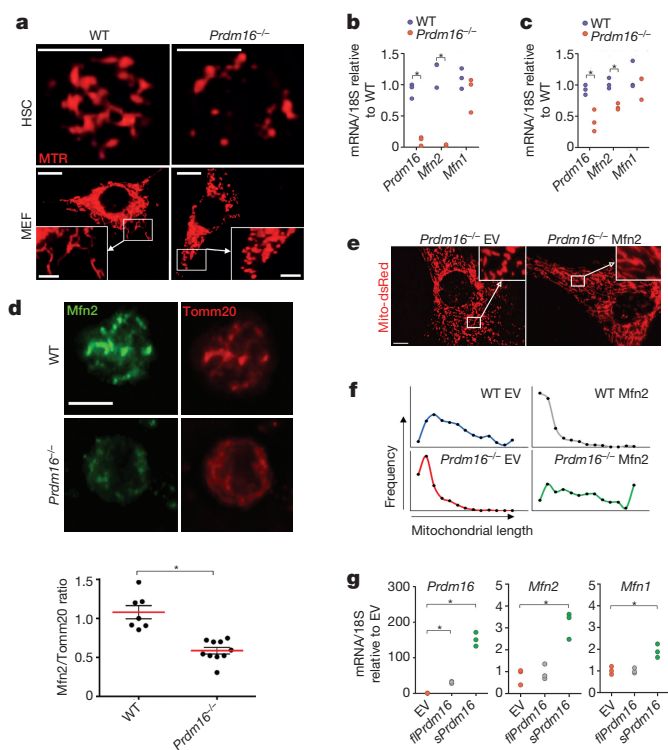


Figure 1 | Prdm16 induces Mfn2. **a**, Mitochondrial morphology of MTR-stained wild-type (WT) and *Prdm16*^{-/-} fetal liver HSCs (Lin⁻Sca1⁺kit⁺CD48⁻CD150⁺Flt3⁻) (scale bars, 5 μ m) and MEFs (scale bars, 20 μ m; inset, 5 μ m). **b**, Relative mRNA expression of *Prdm16*, *Mfn1* and *Mfn2* in WT and *Prdm16*^{-/-} MEFs. *n* = 3 independent experiments; **P* < 0.05; two-tailed Student's *t*-test. **c**, Relative mRNA expression of *Prdm16*, *Mfn1* and *Mfn2* in WT and *Prdm16*^{-/-} adult HSCs. *n* = 3 biological replicates; **P* < 0.05; two-tailed Student's *t*-test. **d**, Representative immunofluorescence for Mfn2 and Tomm20 in WT and *Prdm16*^{-/-} HSCs (scale bar, 5 μ m) and Mfn2 quantification normalized to Tomm20. Bars, mean \pm s.e.m.; *n* \geq 10 fields of cells from two biological replicates; **P* < 0.05; two-tailed Student's *t*-test. **e**, Mitochondrial morphology in *Prdm16*^{-/-} MEFs co-transduced with Mito-dsRed and Mfn2 or control lentivirus (EV, empty vector) (scale bar, 20 μ m). **f**, Mitochondrial length profiles in WT and *Prdm16*^{-/-} MEFs transduced with EV or Mfn2 (\geq 12 cells and \geq 80 mitochondria from 2 biological replicates). Note that Mfn2 overexpression in WT MEFs causes aggregation of mitochondria and apparent shortening, as reported previously³⁰. **g**, Relative mRNA expression in *Prdm16*^{-/-} MEFs of *Prdm16*, *Mfn1* and *Mfn2* 72 h after retroviral expression of *sPrdm16* and *flPrdm16*. *n* = 3 biological replicates; **P* < 0.05; one-way analysis of variance (ANOVA) with Dunnett's post-hoc test.

reach statistical significance (Fig. 3a and Extended Data Fig. 5f). Lentiviral overexpression of *Mfn2* in wild-type HSCs yielded reciprocal results (Extended Data Fig. 6a–i). As these phenotypic analyses and transplantation experiments suggested selective requirement for *Mfn2* in the maintenance of HSCs with extensive lymphoid potential, we performed competitive single HSC transplantation studies to rigorously determine clonal variation in differentiation potential. Although out of >100 recipients too few mice were reconstituted to statistically assess HSC frequency, among recipients with >0.1% donor contribution most *Mfn2*^{fl/fl}-*Vav-Cre* HSCs were myeloid-dominant, whereas most *Mfn2*^{fl/fl} HSCs were balanced or lymphoid 8 weeks after transplantation. In mice that still showed repopulation after 13 weeks, only myeloid-dominant HSCs were detected recipients of *Mfn2*^{fl/fl}-*Vav-Cre*, while most donor HSCs had extensive lymphoid potential in recipients of *Mfn2*^{fl/fl} cells (Fig. 3b). To more accurately determine HSC frequencies we performed limiting dilution experiments²². Among *Mfn2*^{fl/fl}-*Vav-Cre* HSCs overall repopulating HSC frequency was decreased fourfold compared to *Mfn2*^{fl/fl} HSC (Fig. 3c and Extended Data Table 2). The frequency of HSCs capable of >1% long-term lymphoid

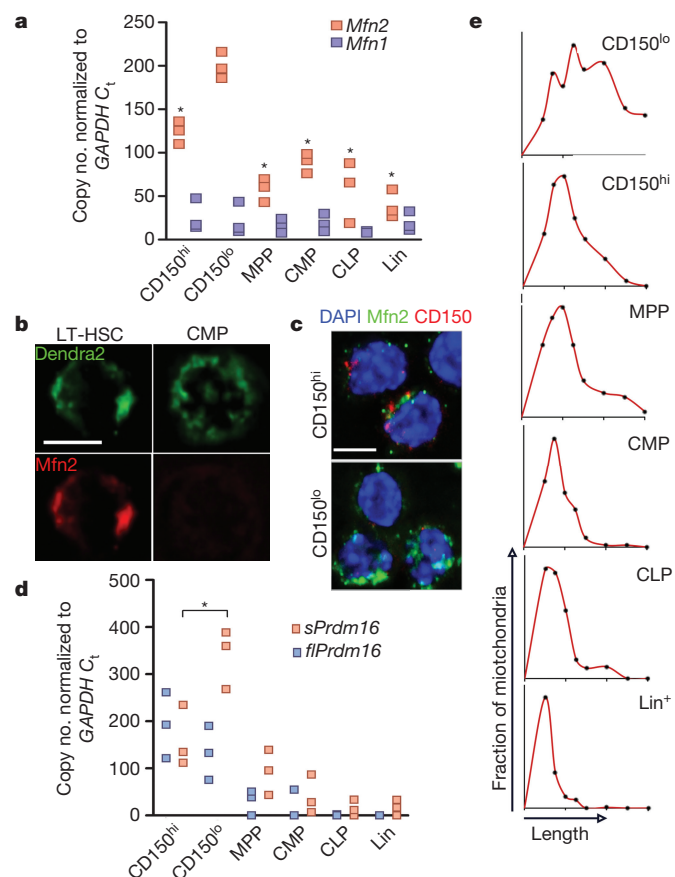
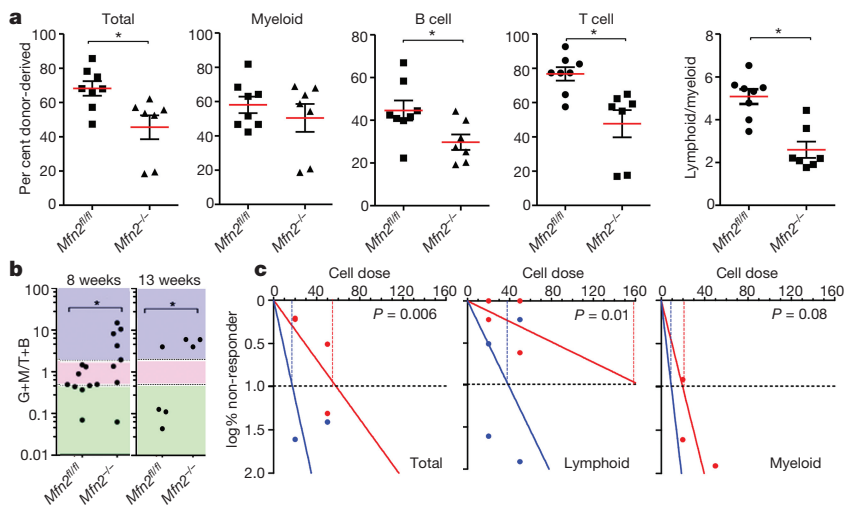


Figure 2 | Mitochondrial morphology and Mfn2 in HSCs. **a**, Expression of *Mfn2* and *Mfn1* mRNA in HSCs, progenitors (MPP, multipotential progenitors, Lin⁻Sca1⁺kit⁺CD150⁺Flt3⁺; CMP, common myeloid progenitors, Lin⁻Sca1⁻Kit⁺; CLP, common lymphoid progenitor, Lin⁻Sca1^{lo}kit^{lo}Flt3⁺IL7R α ⁺ and Lin, lineage⁺ cells). *n* = 3 biological replicates; **P* < 0.05; one-way ANOVA with Dunnett's post-hoc test. **b**, Mfn2 immunofluorescence in Pham-reporter⁺ HSCs and CMPs (Pham mice express a mitochondrially targeted Dendra2 fluorescent protein²⁰) (scale bar, 5 μ m). **c**, Immunofluorescence staining for Mfn2 and CD150 in CD150^{hi} and CD150^{lo} LSKCD48⁻ HSCs (scale bar, 5 μ m). Cell nuclei counterstained with DAPI. **d**, Quantification of *flPrdm16* and *sPrdm16* mRNA in HSCs, progenitors and Lin⁺ cells. *n* = 3 biological replicates; **P* < 0.05; one-way ANOVA with Dunnett's post-hoc test. **e**, Mitochondrial length frequency profile in Pham-reporter⁺ HSCs, progenitors and Lin⁺ cells (*n* \geq 15 fields from three biological replicates).

reconstitution was also approximately fourfold lower. However, the decrease in the frequency of HSCs capable of >1% myeloid reconstitution did not reach statistical significance (Fig. 3c and Extended Data Table 2). Taken together, these results indicate that *Mfn2* is required for the maintenance of HSCs with extensive lymphoid potential.

Next, we identified the mechanism of action of Mfn2. Apoptosis and reactive oxygen species production were similar in *Mfn2*^{fl/fl}-*Vav-Cre* and *Mfn2*^{fl/fl} HSCs, but *Mfn2*^{fl/fl}-*Vav-Cre* HSCs and progenitors displayed increased expression of Grp78, a marker of endoplasmic reticulum stress, which has been shown to be associated with *Mfn2* deletion²³ (Extended Data Fig. 7a–c). *Mfn2*, but not *Mfn1*, also tethers mitochondria to the endoplasmic reticulum, thereby enhancing intracellular calcium buffering⁶. Indeed, intracellular Ca²⁺ was increased in *Mfn2*^{fl/fl}-*Vav-Cre* compared to *Mfn2*^{fl/fl} HSCs (Fig. 4a), in CD150^{hi} compared to CD150^{lo} HSCs (Fig. 4b) and in *Prdm16*^{-/-} compared to wild-type LSK cells (Fig. 4c), but was decreased after lentiviral transduction of wild-type HSCs (Extended Data Fig. 7d). However, Mfn2 did not affect ATP- or SDF1-induced intracellular Ca²⁺ transients (Fig. 4a–c and Extended Data Fig. 7d). At variance with these data



and with a previous report showing only delayed removal of intracellular Ca^{2+} (ref. 6), we found both lower baseline intracellular Ca^{2+} and lower amplitude of ATP-induced calcium flux in *Mfn2*^{-/-} MEFs (Extended Data Fig. 7e). Thus, despite cell type-specific differences in calcium homeostasis, *Mfn2* negatively regulates intracellular Ca^{2+} .

Sustained increase in intracellular Ca^{2+} activates calcineurin, which dephosphorylates four isoforms of Nfat and promotes their translocation to the nucleus²⁴, where they orchestrate multiple processes^{25,26}. We therefore examined whether *Mfn2* inhibits Nfat. As measured by

Figure 3 | Role of *Mfn2* in HSC function. **a**, Donor (CD45.2) chimaerism 15 weeks after competitive transplantation of 2×10^5 *Mfn2*^{fl/fl} or *Mfn2*^{-/-}. *Vav-Cre* (*Mfn2*^{-/-}) adult bone marrow cells together with 2×10^5 CD45.1⁺ competitor bone marrow cells into CD45.1⁺CD45.2⁺ recipients. Plots, mean \pm s.e.m.; $n = 7$ –8 recipients from two independent transplants; $*P < 0.05$; two-tailed Student's *t*-test. **b**, Donor GM/B+T reconstitution ratios 8 (left) and 13 (right) weeks after transplantation of single HSCs from *Mfn2*^{fl/fl} or *Mfn2*^{-/-}. *Vav-Cre* (*Mfn2*^{-/-}) mice. Plots, mean \pm s.e.m.; $n \geq 8$ recipients; $*P < 0.05$; Student's *t*-test. Colours represent myeloid-biased (blue), balanced (pink) and lymphoid dominant (green) as defined in ref. 11. **c**, Limiting dilution assay (see Methods) with *Mfn2*^{fl/fl} or *Mfn2*^{-/-}. *Vav-Cre* (*Mfn2*^{-/-}) adult bone marrow HSCs co-transplanted with CD45.1⁺ competitor bone marrow cells analysed for total, myeloid or lymphoid potential 15 weeks after transplantation. Plots, log% negative responders; $n = 2$ independent experiments with 4–5 recipients each per cell dose; frequencies calculated using limiting dilution analysis Poisson distribution; $P < 0.05$; Pearson's chi-squared test.

immunofluorescence, Nfat1 nuclear localization was decreased after lentiviral transduction of wild-type HSCs (Extended Data Fig. 7f), and increased in *Mfn2*^{fl/fl}-*Vav-Cre* compared to *Mfn2*^{fl/fl} HSCs (Fig. 4d) and in *Mfn2*^{-/-} compared to wild-type MEFs (Extended Data Fig. 7g), which was confirmed by cellular fractionation followed by western blot (Fig. 4e). Consistent with inhibition of Nfat nuclear localization by *Mfn2*, the fraction of nuclear Nfat was also higher in CD150^{hi} compared to CD150^{lo} HSCs (Fig. 4f) and in *Prdm16*^{-/-} compared to wild-type HSCs (Fig. 4g). We next assessed the effect of

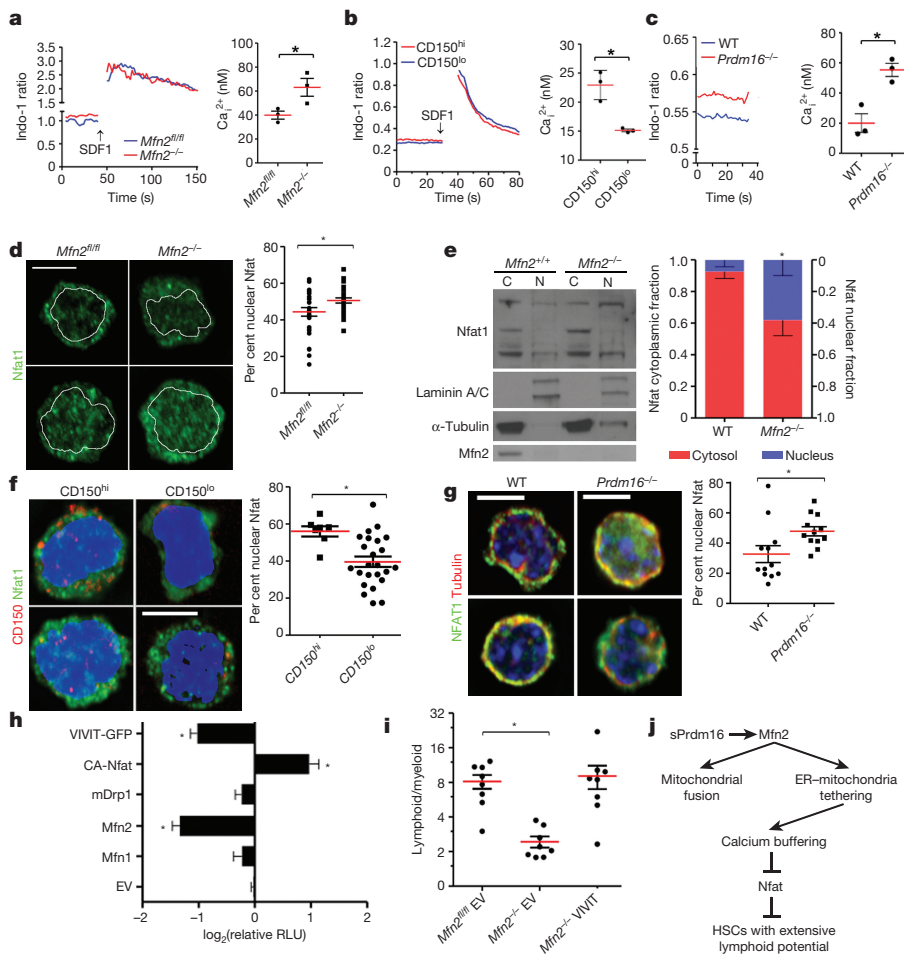


Figure 4 | Mechanism of action of *Mfn2*.

a–c, Calcium flux trace (left) and baseline intracellular Ca^{2+} (right) in *Mfn2*^{fl/fl} or *Mfn2*^{-/-}. *Vav-Cre* (*Mfn2*^{-/-}) HSCs (**a**), CD150^{hi} and CD150^{lo} HSCs (**b**), and WT and *Prdm16*^{-/-} HSCs (**c**). Plots, mean \pm s.e.m.; $n = 3$ biological replicates; $*P < 0.05$; two-tailed Student's *t*-test. **d**, Nfat1 staining (left; scale bar, 5 μm) and percentage of nuclear Nfat1 (right, $*P < 0.05$) in *Mfn2*^{fl/fl} or *Mfn2*^{-/-}. *Vav-Cre* (*Mfn2*^{-/-}) HSCs. Plot, mean \pm s.e.m.; $n \geq 7$ fields of cells from two biological replicates; $*P < 0.05$; two-tailed Student's *t*-test. **e**, Subcellular fractionation followed by western blot for Nfat1 in WT and *Mfn2*^{-/-} MEFs (left, see Supplementary Fig. 1 for full scan) and relative quantification of nuclear and cytoplasmic fraction (right). Bars, mean \pm s.d.; $n = 3$ biological replicates; $*P < 0.05$; two-tailed Student's *t*-test. **f**, **g**, Nfat1 staining (left, scale bars = 5 μm) and percentage of nuclear Nfat1 (right, $*P < 0.05$) in CD150^{hi} and CD150^{lo} HSCs (**f**), and WT and *Prdm16*^{-/-} fetal liver HSCs (**g**). Plots, mean \pm s.e.m.; $n \geq 7$ fields of cells pooled from three biological replicates; $*P < 0.05$; two-tailed Student's *t*-test. **h**, Luciferase activity of 3 \times Nfat gene reporter in 3T3 cells transfected with constructs on y axis (EV, empty vector). Data normalized to β -Gal activity and relative to EV. CA-Nfat, constitutively active Nfat; RLU, relative light units. Bars, mean \pm s.d.; $n = 7$ biological replicates; $*P < 0.05$; one-way ANOVA with Dunnett's post-hoc test. **i**, Lymphoid/myeloid ratio 15 weeks after competitive transplantation of *Mfn2*^{fl/fl} or *Mfn2*^{-/-}. *Vav-Cre* (*Mfn2*^{-/-}) HSCs transduced with IRES-GFP (EV) or VIVIT-GFP. Plots, mean \pm s.e.m.; $n > 8$ recipients pooled from three independent transplants; $*P < 0.05$; one-way ANOVA with Dunnett's post-hoc test. **j**, Schematic representation of the mechanism of action of *Mfn2*.

Mfn2 on Nfat transcriptional activity. Transfection of *Mfn2*^{-/-} MEFs with a Nfat-responsive luciferase reporter revealed increased Nfat transcriptional activity compared to wild-type MEFs (Extended Data Fig. 7h). Conversely, overexpression of *Mfn2* in NIH-3T3 fibroblasts reduced Nfat-responsive luciferase reporter activity to the same extent as VIVIT, a membrane-permeable peptide that specifically inhibits the interaction between calcineurin and Nfat (Fig. 4h)²⁷. However, *Mfn1* and a dominant negative DRP1 mutant (Extended Data Fig. 7i) had no effect (Fig. 4h). Unlike *Mfn2*, *Mfn1* and *Drp1* only regulate mitochondrial fusion and fission. These findings therefore suggest that Mfn2 inhibits Nfat through its endoplasmic reticulum-mitochondria tethering activity rather than through mitochondrial fusion. Finally, and consistent with *Mfn2* induction by *sPrdm16*, lentiviral transduction of *Prdm16*^{-/-} HSCs with *sPrdm16*, but not with *flPrdm16*, reduced Nfat nuclear localization (Extended Data Fig. 7j), while Nfat transcriptional activation was increased in *Prdm16*^{-/-} MEFs and normalized by transfection of either *sPrdm16* or *Mfn2*, but not *flPrdm16* (Extended Data Fig. 7k). We conclude that *sPrdm16*-mediated induction of *Mfn2* inhibits Nfat activity.

To directly examine the role of Nfat downstream of *Mfn2*, we inhibited its function in HSCs. Culture of wild-type HSCs in the presence of VIVIT peptide increased expression of CD150^{lo} HSC-associated lymphoid commitment markers, *Il7r* and *Sox4* (ref. 18; Extended Data Fig. 8a), and increased lymphoid/myeloid repopulation ratio after subsequent competitive transplantation (Extended Data Fig. 8b). Most importantly, lentiviral VIVIT-GFP²⁷ transduction fully rescued the long-term lymphoid reconstitution defect of *Mfn2*^{fl/jl}-*Vav-Cre* HSCs (Fig. 4i) and, similar to *Mfn2* transduction, increased lymphoid repopulation of wild-type HSCs (Extended Data Fig. 8c). Nfat inhibition is therefore the prime mechanism downstream of *Mfn2*.

The observation that *Mfn2*, induced by *sPrdm16*, maintains HSCs with extensive lymphoid potential by negatively regulating calcineurin/Nfat activity through enhanced intracellular Ca²⁺ buffering (Fig. 4l) identifies such HSCs, which decline with age^{28,29}, as a mechanistically defined subset and provides a mechanism underpinning clonal heterogeneity in HSCs. These findings may lead to the design of approaches to bias HSC differentiation into desired lineages after transplantation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 January; accepted 30 November 2015.

Published online 20 January 2016.

- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Simsek, T. et al. The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell* **7**, 380–390 (2010).
- Takubo, K. et al. Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. *Cell Stem Cell* **12**, 49–61 (2013).
- Aguilo, F. et al. *Prdm16* is a physiologic regulator of hematopoietic stem cells. *Blood* **117**, 5057–5066 (2011).
- Chukov, S., Levi, B. P., Smith, M. L. & Morrison, S. J. *Prdm16* promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. *Nature Cell Biol.* **12**, 999–1006 (2010).
- de Brito, O. M. & Scorrano, L. Mitofusin 2 tethers endoplasmic reticulum to mitochondria. *Nature* **456**, 605–610 (2008).
- Rizzuto, R., De Stefani, D., Raffaello, A. & Mammucari, C. Mitochondria as sensors and regulators of calcium signalling. *Nature Rev. Mol. Cell Biol.* **13**, 566–578 (2012).
- Müller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B. & Sieburg, H. B. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood* **100**, 1302–1309 (2002).

- Sanjuan-Pla, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
- Müller-Sieburg, C. E. & Sieburg, H. B. Clonal diversity of the stem cell compartment. *Curr. Opin. Hematol.* **13**, 243–248 (2006).
- Dykstra, B. et al. Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* **1**, 218–229 (2007).
- Chan, D. C. Fusion and fission: interlinked processes critical for mitochondrial health. *Annu. Rev. Genet.* **46**, 265–287 (2012).
- Youle, R. J. & van der Bliek, A. M. Mitochondrial fission, fusion, and stress. *Science* **337**, 1062–1065 (2012).
- Tanaka, A. & Youle, R. J. A chemical inhibitor of DRP1 uncouples mitochondrial fission and apoptosis. *Mol. Cell* **29**, 409–410 (2008).
- Shing, D. C. et al. Overexpression of sPRDM16 coupled with loss of p53 induces myeloid leukemias in mice. *J. Clin. Invest.* **117**, 3696–3707 (2007).
- Yoshida, M. et al. Aberrant expression of the *MEL1S* gene identified in association with hypomethylation in adult T-cell leukemia cells. *Blood* **103**, 2753–2760 (2004).
- Sorianello, E. et al. The promoter activity of human *Mfn2* depends on Sp1 in vascular smooth muscle cells. *Cardiovasc. Res.* **94**, 38–47 (2012).
- Beeraman, I. et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc. Natl Acad. Sci. USA* **107**, 5465–5470 (2010).
- Weksberg, D. C., Chambers, S. M., Boles, N. C. & Goodell, M. A. CD150- side population cells represent a functionally distinct population of long-term hematopoietic stem cells. *Blood* **111**, 2444–2451 (2008).
- Pham, A. H., McCaffery, J. M. & Chan, D. C. Mouse lines with photo-activatable mitochondria to study mitochondrial dynamics. *Genesis* **50**, 833–843 (2012).
- Chen, H., McCaffery, J. M. & Chan, D. C. Mitochondrial fusion protects against neurodegeneration in the cerebellum. *Cell* **130**, 548–562 (2007).
- Szilvassy, S. J., Humphries, R. K., Lansdorp, P. M., Eaves, A. C. & Eaves, C. J. Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy. *Proc. Natl Acad. Sci. USA* **87**, 8736–8740 (1990).
- Ngoh, G. A., Papanicolaou, K. N. & Walsh, K. Loss of mitofusin 2 promotes endoplasmic reticulum stress. *J. Biol. Chem.* **287**, 20321–20332 (2012).
- Li, H., Rao, A. & Hogan, P. G. Interaction of calcineurin with substrates and targeting proteins. *Trends Cell Biol.* **21**, 91–103 (2011).
- Müller, M. R. et al. Requirement for balanced Ca/NFAT signaling in hematopoietic and embryonic development. *Proc. Natl Acad. Sci. USA* **106**, 7034–7039 (2009).
- Wu, H., Peisley, A., Graef, I. A. & Crabtree, G. R. NFAT signaling and the invention of vertebrates. *Trends Cell Biol.* **17**, 251–260 (2007).
- Aramburu, J. et al. Affinity-driven peptide selection of an NFAT inhibitor more selective than cyclosporin A. *Science* **285**, 2129–2133 (1999).
- Benz, C. et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. *Cell Stem Cell* **10**, 273–283 (2012).
- Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M. & de Haan, G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J. Exp. Med.* **208**, 2691–2703 (2011).
- Chen, H. et al. Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J. Cell Biol.* **160**, 189–200 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements The pHR-IRES-GFP SIN vector was a gift from R. Creusot. *Mfn2*^{-/-} MEFs were a gift from D. Chan. We thank S.-H. Ho and E. A. Schon for assistance and discussion. This work was supported by grant NIH RO1 CA167286 and RO1 AG029262 (H.-W.S.), the Druckenmiller Fellowship from the New York Stem Cell Foundation (L.L.L.), the Ruth L. Kirschstein F31 CA196045 fellowship (D.J.C.), the NIH National Center for Research Resources award 1S10RR027050-01 and the NIH Office of the Director award 1S10OD020056-01.

Author Contributions L.L.L. designed and performed most experiments, contributed to the concept, and co-wrote the manuscript with H.-W.S.; M.J.A. generated the data in Fig. 4e and Extended Data Fig. 8a, and assisted in Figs 2, 4i and Extended Data Figs 6–8. M.M. cloned the DNDp1 construct and verified its function (Extended Data Fig. 7d). D.J.C. generated the *Mfn2* promoter construct (Extended Data Fig. 3c). H.-W.S. provided concept and guidance, and co-wrote the paper with L.L.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.-W.S. (hs2680@columbia.edu).

METHODS

Animals. C57BL/6J mice (CD45.2) and B6.SJL-Ptprca^{Pep3b/BoyJ} (CD45.1) were purchased from The Jackson Laboratory (Bar Harbour, ME). *Prdm16*^{Gt(OST67423)Lex} knockout mice³¹ were obtained from Lexicon Genetics. Conditional MitoDendra2 transgenic (Pham) mice³² (B6;129S-Gt(ROSA)26Sor^{tm1(CAG-COX8A/Dendra2)Dcc/J}) and E2A-Cre mice³³ (B6.FVB-Tg(EIIa-cre)C5379Lmgd/J) were purchased from Jackson Laboratory. Pham mice contain a mitochondrially targeted Dendra preceded by a stoplox sequence in the Rosa locus. These mice were crossed with E2A-Cre mice to effect ubiquitous induction of the MitoDendra2 reporter. Conditional *Mfn2* knockout mice³⁴ (B6/129SF1^{Mfn2tm3Dcc/Mmucd}) were obtained from MMRRC and crossed to Vav-Cre transgenic mice³⁵ (B6.Cg-Tg(Vav1-Cre)A2Kio/J) to obtain a homozygous floxed allele *Mfn2* allele which generated a B6.Cg-Tg(Vav1-Cre)A2Kio/J;B6/126SF1^{Mfn2tm3Dcc/Mmucd} mixed mouse strain. All mouse strains were rederived by *in vitro* fertilization at the Jackson Laboratory. Animals were housed in a specific pathogen-free facility. Experiments and animal care were performed in accordance with the Columbia University Institutional Animal Care and Use Committee. All mice were used at age 8–12 weeks, except in experiments that involved fetal liver cells, when E14.4 embryos were used. Both sexes were used for experiments. Results were analysed in non-blinded fashion. In all experiments, randomly chosen wild type and littermates were used.

MEF isolation and cell lines. MEFs were established from approximately 14.5 days post coitum embryos as previously described³⁶ from *Prdm16*^{+/-} breeder pairs. Briefly, dissected embryo trunks were minced into 1–2 mm fragments, resuspended in 3 ml 0.25% trypsin/EDTA (Gibco, Carlsbad, CA) and passed 20–30 times through a 16 gauge needle. Cell suspensions were incubated at 37 °C for 1 h with frequent agitation. Erythrocytes were lysed with ACK buffer, washed and cells were plated for 3 h in 10% FBS/DMEM. Cells remaining in suspension were aspirated and adherent cells were cultured with fresh media. MEFs were passaged 1:3 every 3 days and cells between passage 2 and 5 were used for all experiments. 293 cells and NIH-3T3 cells were purchased from ATCC (Manassas, VA) and sub-cultured in 10% FBS/DMEM or 10% calf serum/DMEM, respectively. WT and *Mfn2*^{-/-} MEFs were a kind gift from E. Schon (Columbia University). All lines are tested yearly for mycoplasma contamination and found negative.

Plasmids. *Prdm16* constructs were generated by subcloning the murine full length (*flPrdm16*) or truncated (*sPrdm16*) cDNA into the XhoI/EcoRI sites of the pMSCV-IRES-GFP retroviral expression plasmid. The Mito-dsRed construct was purchased from Addgene (Cambridge, MA) (plasmid 11151). *Mfn2* constructs were generated by subcloning the murine *Mfn2* cDNA into the EcoRI/BamHI sites of the pLVX-EF1α-IRES-GFP or pLVX-EF1α-IRES-mCherry lentiviral expression plasmid (Clontech). The pGreenFire-Nfat and pGreenFire-CMV gene reporter constructs were purchased from System Biosciences (San Jose, CA) and contained three canonical Nfat response elements (5'-GGAAAN-3') driving the expression of copGFP and luciferase reporters. The DNDrp1-pcDNA3.1 construct was purchased from Addgene (#45161) and subcloned using the BamHI/EcoRI restriction sites into the pLVX-IRES-GFP vector. Lentiviral 2nd generation packaging construct ΔR8.2 (8455) and pDM2.6 (12259) were purchased from Addgene. The -950/+22 murine *MFN2* promoter was constructed by PCR amplification of the RP23-458J18 BAC clone (CHORI, Oakland, CA) and subcloned into the pGL4 luciferase reporter vector (Promega, Madison, WI). All cloning was carried out using KOD hot-start polymerase (Novagen, Billerica, MA) and subcloned for screening and sequencing into the pCR2.1 shuttle vector (Invitrogen, Carlsbad, CA).

FACS sorting and analysis. For peripheral blood analyses, erythrocytes were lysed twice with ACK lysis buffer and nucleated cells were stained with antibody cocktail (Supplementary Table 1) in FACS buffer for 15 min on ice, washed and analysed on a BD FACSCantoII flow cytometer (Becton Dickinson, Mountain View, CA). For bone marrow analyses, cells were isolated using the crushing method and erythrocytes were lysed with ACK lysis buffer followed by 40 μm filtration. Bone marrow cells were stained with antibody cocktail in FACS buffer for 30 min on ice, washed and analysed on a BD LSRII flow cytometer (Becton Dickinson, Mountain View, CA). Dead cells were excluded from analyses by gating out 7AAD-positive cells. To isolate purified haematopoietic populations, bone marrow cells were isolated, stained and sorted using a BD Influx cell sorter (Becton Dickinson, Mountain View, CA) into complete media. Data were analysed using FlowJo9.6 (TreeStar Inc., Ashland, OR).

Haematopoietic stem cell transplantation. *Mfn2*^{fl/fl}-Vav-Cre fetal liver cells, bone marrow cells or purified LT-HSCs (Lin⁻cKit⁺Sca1⁺CD48⁻Flt3⁻CD150⁺) were transplanted into lethally irradiated (two doses of 478 cGy over 3 h using a Rad Source RS-2000 X-ray irradiator (Brentwood, TN)) recipients together with 2 × 10⁵ competitor cells. As *Mfn2*^{fl/fl}-Vav-Cre mice were not fully backcrossed onto the C57BL/6 background, recipient mice and competitor bone marrow cells were from the B6.Cg-Tg(Vav1-Cre)A2Kio/J;B6/126SF1^{Mfn2tm3Dcc/Mmucd} mixed

background mouse strain crossed to B6.SJL-Ptprca^{Pep3b/BoyJ} (CD45.1) to generate a CD45.1⁺CD45.2⁺ mixed background mouse. Competitor cells were T-cell depleted using MACS beads. For all competitive transplantation experiments, at least two independent transplants, each with at least 4 recipients per condition of genotype were performed, and result of all recipients pooled for statistical analysis. Power calculation was based on results of the first experiment. In limiting dilution assays, cohorts of recipients received 20 or 50 HSCs together with 2 × 10⁵ competitor cells, allowing calculation of HSC frequency based on the number of non-repopulated mice (<1% donor contribution) using Poisson statistics 15 weeks after reconstitution. For *Mfn2* KO single cell transplantation, LT-HSCs were sorted directly into complete media (StemPro34, 100 ng ml⁻¹ SCF, 100 ng ml⁻¹ TPO, 50 ng ml⁻¹ IL-6) and single cells were visually confirmed. Positive single cell wells were combined with 2 × 10⁵ CD45.1 competitor bone marrow cells and transplanted into lethally irradiated CD45.1 recipient mice. Recipients showing ≥ 0.1% CD45.2 donor contribution were considered positive and GM/(B+T) ratios were calculated as previously described for characterizing heterogeneous HSC phenotypes³⁷. In transplantations using WT or *Prdm16*^{-/-} HSCs (Lin⁻cKit⁺Sca1⁺CD48⁻Flt3⁻CD150⁺) B6.CD45.2 cells were mixed with 2 × 10⁵ freshly isolated B6.CD45.1 bone marrow cells and injected via tail vein into lethally irradiated (two doses of 478 cGy over 3 h using a Rad Source RS-2000 X-ray irradiator (Brentwood, TN)) B6.CD45.1⁺CD45.2⁺ F1 hybrid recipients. After 8 to 15 weeks, peripheral blood (PB) and bone marrow were analysed.

Lentivirus production, transduction and integration verification. Lentiviral particles were produced by seeding 293 cells at 7 × 10⁵ per cm², or PlatE cells (Cell Biolabs, San Diego, CA), in Ultra Culture serum-free media (Lonza, Basel, Switzerland) overnight followed by transfection of each packaging and expression construct (1:1:1) using Trans-It 293 (Mirus, Madison, WI) for 2 h. Media were pooled after 36–48 h, clarified and concentrated by ultracentrifugation (100,000g), resuspended in StemPro-34 media and stored at -80 °C. Virus titre was calculated from transduction of NIH-3T3 fibroblasts serial dilutions of the viral preparation. Sorted LT-HSCs were transduced with ≥ 150 MOI lentivirus particles in the presence of 6 μg ml⁻¹ polybrene (Sigma) and spun at 900g for 20 min at 20 °C. Supernatant was aspirated and replaced with complete media and cultured overnight. Transduction efficiency of cells was confirmed after 24 h. To assess proviral copy number 15 weeks post-transplantation *in vivo*, splenocytes were harvested and sorted into donor (CD45.2) or competitor (CD45.1) populations and gDNA was isolated as previously described³⁸. Amplification of the proviral WPRE region was achieved using SYBR Green qPCR assay using the primer pair WPREFor: 5'-CCGTTGTCAGGCAACGTG-3' and WPRERev: 5'-AGCTGACAGGTGGTGGCAAT-3'. Quantification of proviral copies was derived from the linear regression of serial dilutions of viral vector and normalized to input cell number.

Quantitative RT-PCR. Sorted or cultured cell populations (2–5 × 10³ cells) were lysed in TRIzol LS reagent (Invitrogen, Carlsbad, CA) and RNA was isolated according to manufacturer's instructions. cDNA was synthesized using Superscript III Reverse Transcriptase (Invitrogen) and target CT values were determined using inventoried TaqMan probes (Applied Biosystems, Carlsbad, CA, see Supplementary Table 2) spanning exon/exon boundaries and detected using a Viia7 Real Time PCR System (Applied Biosystems). Relative quantification was calculated using the ΔΔC_T method. To estimate relative copy number of *Mfn1* and *Mfn2* transcripts (Fig. 4a), copy numbers were derived from the linear regression of serial dilutions of respective cDNA plasmids and normalized to *GAPDH-VIC* values. To estimate relative copy number of *flPrdm16* transcripts (Fig. 4d), a probe was designed to span the SET methyltransferase domain of *Prdm16* (exon2/3 junction) and copy number was derived from the linear regression of serial dilutions of respective cDNA plasmids. Another probe (exon 14/15 junction) was used to quantify total *Prdm16* copy numbers derived from the linear regression of serial dilutions of respective cDNA plasmids. The values derived from total *Prdm16* probe was subtracted from *flPrdm16*-specific probe to determine *sPrdm16* transcript quantity. All values were normalized to relative multiplexed *GAPDH-VIC* values.

LT-HSC culture. Culture of sorted LT-HSCs was carried out using StemPro34 media (Invitrogen) supplemented with 10 mM HEPES and 50 ng ml⁻¹ of recombinant murine SCF, TPO, IL-6 (Peprotech, Rocky Hill, NJ) and cultured in 5% O₂ at 37 °C. In some experiments, LT-HSCs were cultured in the presence of 500 ng ml⁻¹ VIVIT (Millipore, Billerica, MA) or 30 μM mDivi1 (MolPort, Riga, Latvia).

Mitochondrial PEG-1500 fusion assay. To demonstrate a mitochondrial fusion activity, cell fusion experiments were performed using MEFs as previously described³⁷. Briefly, BacMam baculovirus constructs (Invitrogen) expressing the signalling peptide from cytochrome *c* fused to either GFP or RFP were transduced separately into MEF cells. Sorted GFP⁺ and RFP⁺ MEFs were co-cultured for 24 h

and plasma membranes were fused using PEG-1500 (Roche, Basel, Switzerland). Fused cells were cultured in DMEM containing cycloheximide (Sigma, St. Louis, MO) for 4 h and analysed for colocalization of mitochondrial labels.

Chromatin immunoprecipitation (ChIP). Early passage *Prdm16*^{-/-} MEFs were transduced with 10 MOI retrovirus for 72 h and fixed with 4% paraformaldehyde for 10 min. Protein lysates were isolated and chromatin immunoprecipitation was carried out using the ChIP-IT Express Enzyme kit (Active Motif, Carlsbad, CA). Antibodies used for ChIP include anti-Flag and anti-TF2D. Primer probes were designed to span regions of the Mfn2 promoter previously shown to regulate Mfn2 transcriptional activity (see Supplementary Table 3)³⁹. Quantification of precipitated Mfn2 promoter regions were derived from the linear regression of serial dilutions of bone marrow genomic DNA, normalized to input DNA concentration and quantifiable IgG detection was subtracted from sample values.

Indo-1-AM calcium flux. Bone marrow was freshly isolated and lineage depleted with the MACS Lineage Depletion Kit (Miltenyi Biotec, San Diego, CA). Cells were cultured for 30 min in complete medium supplemented with 1 μ M Indo-1 prepared as stock supplemented with Pluronic-F127 and incubated at 37 °C for 30 min. Cells were washed and stained for surface markers for 15 min, washed and allowed to rest in for 15 min PBS in PBS with Ca²⁺. FACS tubes were run at 37 °C in the sample port of the LSRII flow cytometer equipped with a 355 nm excitation laser. Events were collected for 40 s before incubation with 25 μ M ATP or 1 μ M SDF1 to induce calcium transients. The average ratio, *R*, of bound/free Indo-1 (405 nm/485 nm emission) before simulation was used to determine baseline values. Identical samples were equilibrated in 10 mM EGTA PBS without Ca²⁺ to determine *R*_{min} or stimulated with 1 μ M ionomycin to determine *R*_{max}. The Indo-1 dissociation constant (*K*_d) was assumed to be 237 nM at 37 °C based on previous studies⁴⁰. The following equation was then used to relate Indo-1 intensity ratios to [Ca²⁺]_i levels;

$$[\text{Ca}^{2+}] = K_d \cdot \frac{(R - R_{\min})}{(R_{\max} - R)}$$

Immunofluorescence. Sorted or cultured haematopoietic populations (2–5 × 10³ cells) were collected in complete media and plated on onto MicroWell 96-well glass-bottom plates (Thermo, Waltham, MA) coated with 1 μ g ml⁻¹ poly-D-lysine. Cells were allowed to adhere for 10 min and fixed with 4% PFA for 15 min. Cells were then permeabilized with 0.1% TritonX-100/PBS for 5 min and blocked with 2% BSA/PBS for 1 h at 4 °C. Cells were incubated with anti-Nfat1 (1:100), anti-Mfn2 (1:200), anti-tubulin (1:200), anti CD150-APC (1:100) or anti-Flag (1:250) (see Supplementary Table 1) overnight, washed and incubated with AlexaFluor secondary antibodies (Invitrogen) for 1 h. Cell nuclei were counterstained with DAPI and mounted with fluorescent mounting media (Vector Labs, Burlingame, CA). Confocal images were acquired with a Zeiss LSM 700 confocal microscope or a Leica DMI 6000B and images were deconvoluted and processed with Leica AF6000 software package.

Gene reporter assays. NIH-3T3, WT or *Mfn2*^{-/-} MEF cells were plated at 2 × 10⁴ cells per cm² in triplicate overnight and transfected with 500 ng of pGF-Nfat, pGF-CMV or -950/+22 Mfn2-pGL4 reporter construct, 500 ng of cDNA plasmids as indicated and 500 ng of either pSV- β Gal or pLVX-IRES-mCherry plasmids with Lipofectamine 3000 according to manufacturer's instructions for 24 or 48 h. Cells were lysed in reporter lysis buffer (Promega, Madison, WI) and analysed for luciferase activity using BrightGlo luciferase (Promega) and detected on a Synergy H2 plate reader (BioTek, Winooski, VT). To visualize β Gal activity, cell lysate was incubated in Buffer Z (1 mg ml⁻¹ ONPG, 0.1 M phosphate, pH 7.5, 10 mM KCl, 1 mM β ME, 1 mM MgSO₄) at 37 °C for 1 h. Absorbance values were measured at 405 nm and used to normalize for transfection efficiency. In WT and *Prdm16*^{-/-} MEFs, gene reporter luciferase values were normalized to mCherry excitation values.

Western blot. For total cell lysate experiments, MEF cultures were lysed in RIPA buffer, 50 mM Tris pH 7.5, 137 mM NaCl, 0.1% SDS, 0.5% deoxycholate and protease inhibitors (Roche). For subcellular fractionation studies, cells were scraped, washed in PBS. Cell pellets were lysed in 5 × packed cell volume (pcv) Buffer A for 10 min on ice and vortexed for 15 s in the presence of 1/10 volume 3% NP-40. Plasma membrane lysis was verified by trypan blue staining. Lysate was spun at

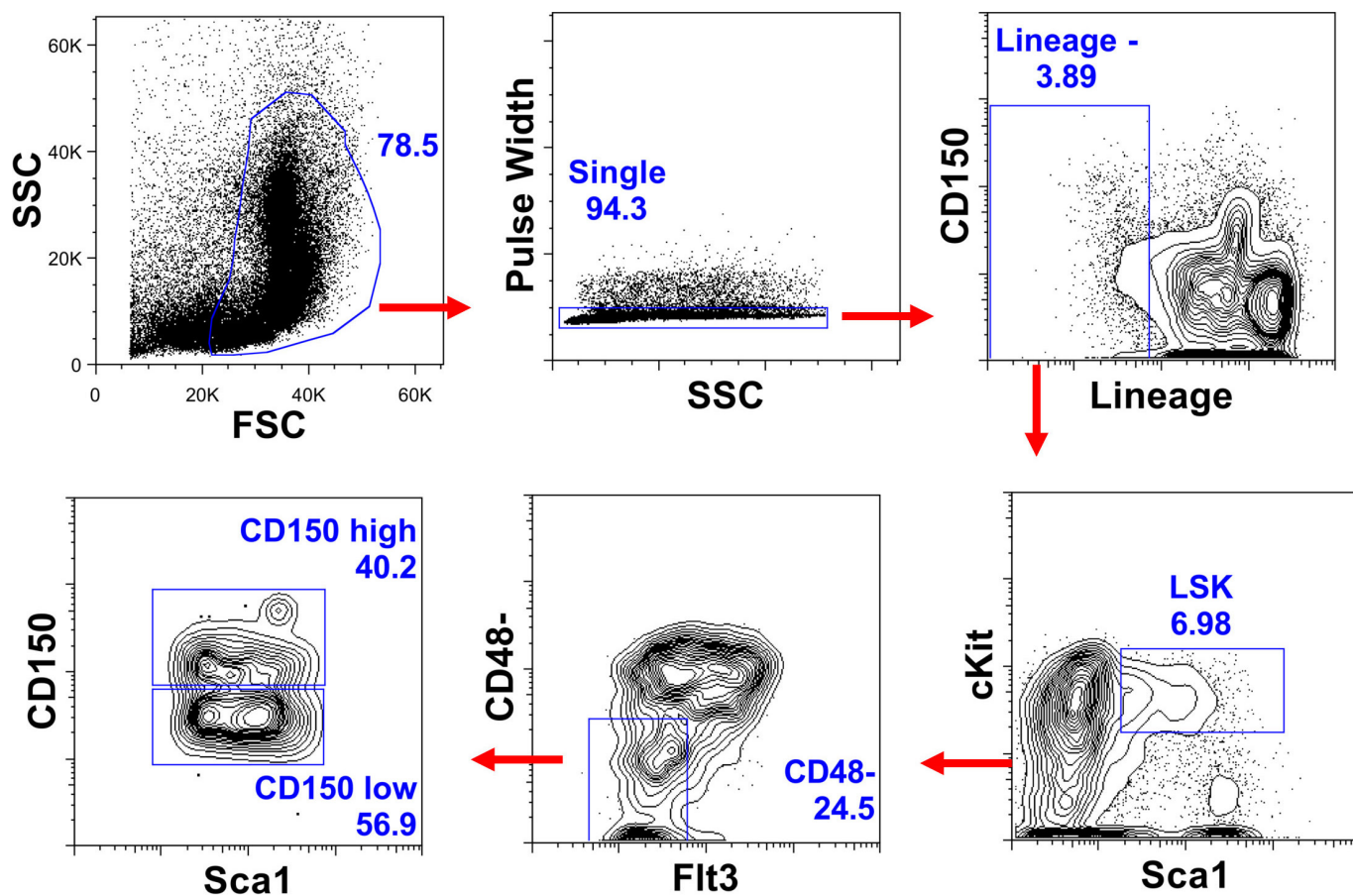
15,000g for 10 min at 4 °C and the cytoplasmic fraction was saved. The remaining nuclear pellet was resuspended in 2.5 × pcv Buffer C and incubated at 4 °C for 1 h with rotation and spun at 15,000g for 10 min. The nuclear fraction was diluted with 2.5 × volume of Nuclear Diluent Buffer and stored at -80 °C. To achieve even fractionation loading, equivalent percentages of nuclear and cytoplasmic fractions were loaded on each gel. All protein samples were denatured in 4 × sample buffer at 95 °C and loaded onto 4–12% Bis-Tris SDS-PAGE gradient gels (Invitrogen). Gels were transferred onto 0.22 μ m nitrocellulose membrane and stained with Ruby Red (Molecular Probes, Carlsbad, CA) to confirm transfer. Membranes were blocked with 3% non-fat milk or BSA in 0.1% Tween-20/TBS and incubated with anti-Mfn2 (1:200), anti- β Gal (1:1,000), anti-Nfat1 (1:250), anti-tubulin (1:1,000), anti-lamin A/C (1:500) and anti- β -actin (1:5,000) overnight (see Supplementary Table 1). Membranes were washed, incubated with HRP-conjugated secondary antibodies and exposed to X-ray film (Denville) after incubation with Super Signal West Femto ECL reagent (Pierce).

Image quantification. For mitochondrial length measurements, confocal or deconvoluted z-stacks were collected and projected as a z-project in ImageJ (NIH, Bethesda, MD). Individual mitochondria were manually traced, binned into length categories and expressed as percent of cellular mitochondria. The mean \pm s.e.m. number of mitochondria falling into each length category collected from ≥ 15 fields (30–50 cells) are expressed. For Nfat nuclear localization quantification, confocal or deconvoluted z-stacks were collected and a 1- μ m section in the centre of the cell was projected as a z-project in ImageJ. Nuclear boundaries were constructed using DAPI staining. The ratio of staining within the nuclear boundary to total staining was expressed as percent of Nfat signal. The mean \pm s.e.m. for ≥ 10 fields (20–40 cells) are expressed. For immunofluorescence intensity measurements, confocal or deconvoluted z-stacks were collected and projected as a z-project in ImageJ. Thresholds were set based on IgG-stained negative control cells and the integrated density value of each signal per cell was recorded. The mean \pm s.e.m. for ≥ 15 fields (30–50 cells) are expressed.

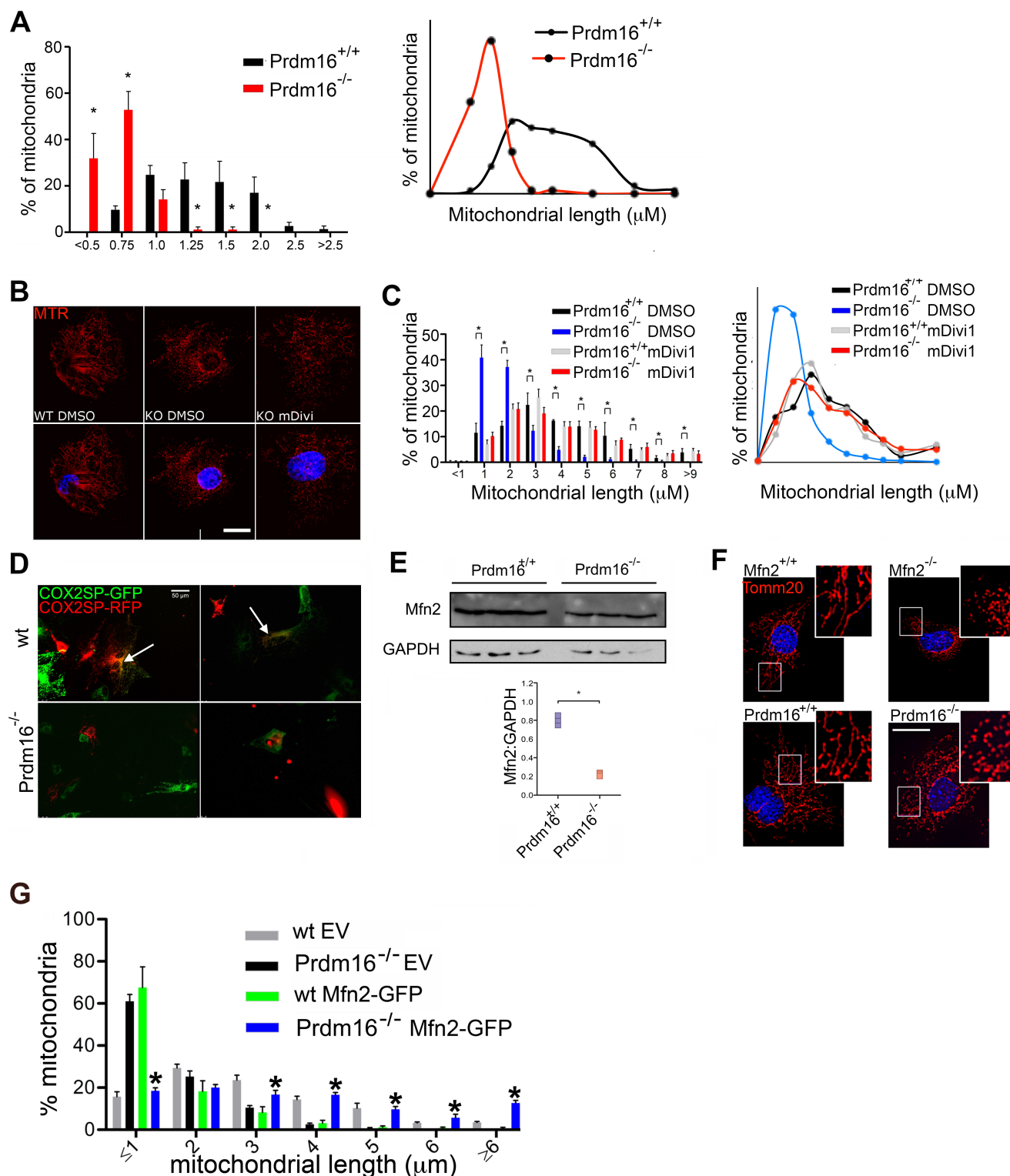
Statistics. For statistical analysis between two groups, the unpaired Student's *t*-test was used. When more than two groups were compared, one-way ANOVA was used. Results are expressed as mean \pm s.e.m. The Bonferroni and Dunnett multiple comparison tests were used for post-hoc analysis to determine statistical significance between multiple groups. All statistics were calculated using Prism5 (GraphPad, La Jolla, CA) software. Differences among group means were considered significant when the probability value, *P*, was less than 0.05. Sample size (*n*) always represents biological replicates. Cochran test was used for exclusion of outliers.

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

- Zambrowicz, B. P. *et al.* Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. *Proc. Natl Acad. Sci. USA* **100**, 14109–14114 (2003).
- Pham, A. H., McCaffery, J. M. & Chan, D. C. Mouse lines with photo-activatable mitochondria to study mitochondrial dynamics. *Genesis* **50**, 833–843 (2012).
- Lakso, M. *et al.* Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl Acad. Sci. USA* **93**, 5860–5865 (1996).
- Chen, H., McCaffery, J. M. & Chan, D. C. Mitochondrial fusion protects against neurodegeneration in the cerebellum. *Cell* **130**, 548–562 (2007).
- Shimshak, D. R. *et al.* Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* **32**, 19–26 (2002).
- Ausubel, F. M. *Current Protocols in Molecular Biology* (Wiley, 1987).
- Benz, C. *et al.* Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. *Cell Stem Cell* **10**, 273–283 (2012).
- Chen, H. *et al.* Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J. Cell Biol.* **160**, 189–200 (2003).
- Sorianello, E. *et al.* The promoter activity of human Mfn2 depends on Sp1 in vascular smooth muscle cells. *Cardiovasc. Res.* **94**, 38–47 (2012).
- Gryniewicz, G., Poenie, M. & Tsien, R. Y. A new generation of Ca²⁺ indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).

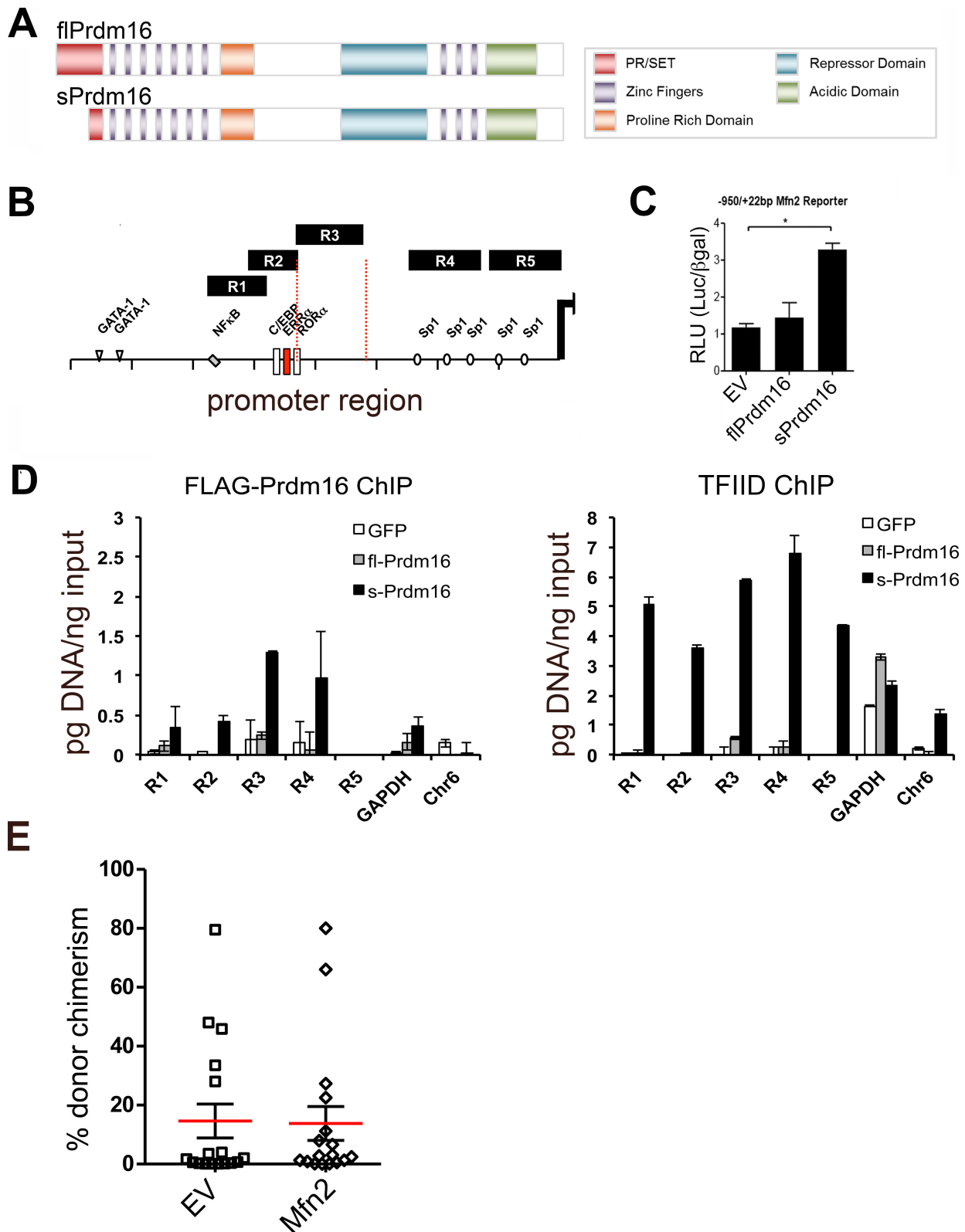


Extended Data Figure 1 | Representative sort gates for isolation of HSCs. Flow cytometric plots showing the gates used to isolate HSCs ($\text{Lin}^- \text{Sca1}^+ \text{kit}^+ \text{Flt3}^- \text{CD48}^-$) and CD150^{hi} and CD150^{lo} HSCs.



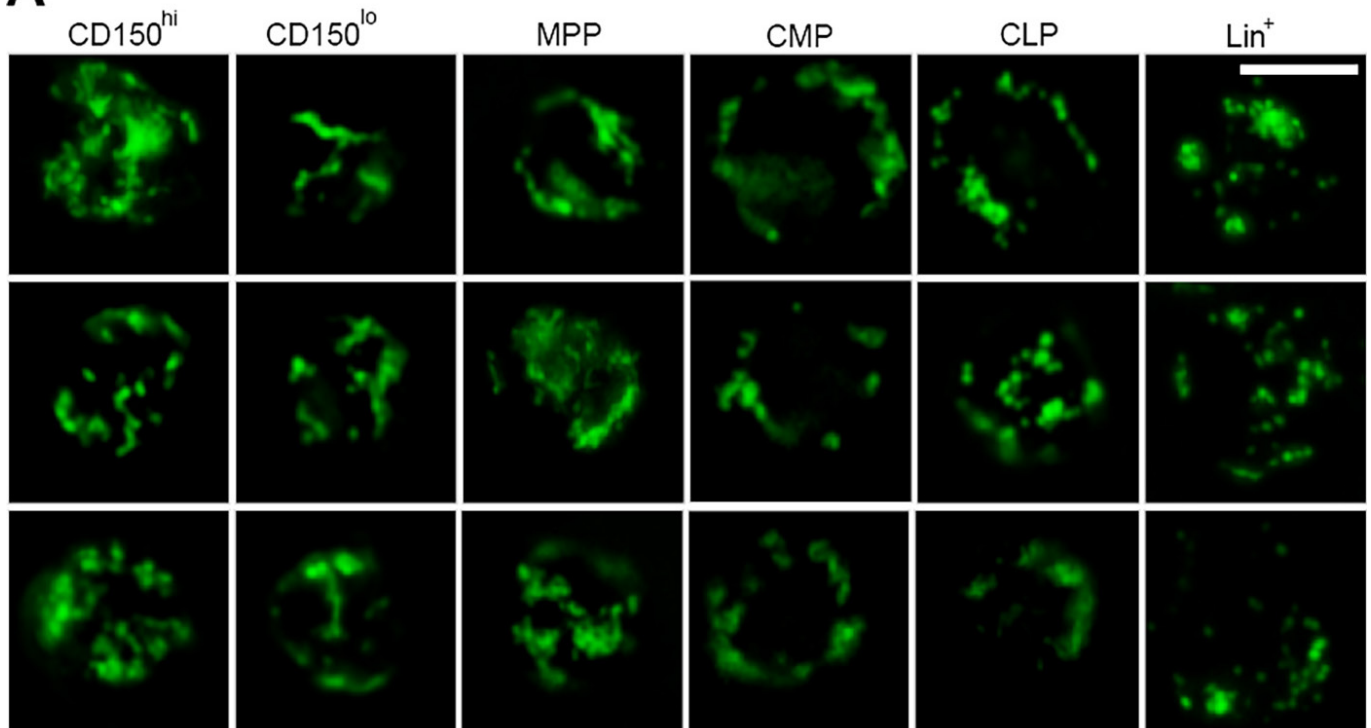
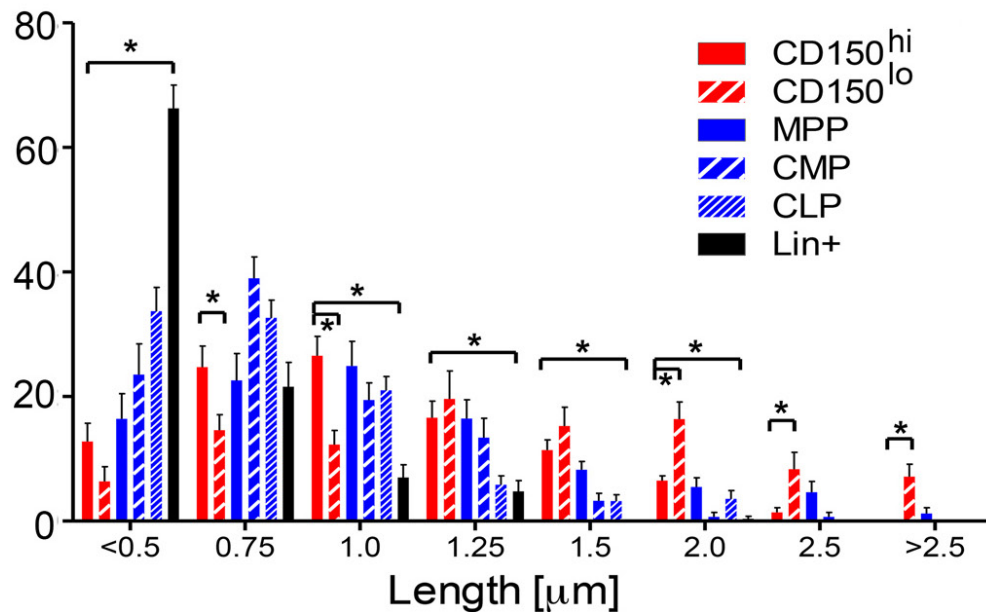
Extended Data Figure 2 | Mitochondrial dynamics in *Prdm16*^{-/-} MEFs. **a**, Frequency distribution (left) and frequency distribution profile (right) of mitochondrial length in WT or *Prdm16*^{-/-} fetal HSCs. Bars, mean ± s.e.m.; *n* ≥ 20 fields of cells from three biological replicates; **P* < 0.05 within length bins; two-tailed Student's *t*-test within length bins. **b**, Mitochondrial morphology in *Prdm16*^{-/-} MEFs treated for 24 h with the Drp1 inhibitor, mDivi (30 μM), or vehicle is shown for comparison, Mitotracker Red staining, scale bar 20 μm. **c**, Frequency distribution (left) and frequency distribution profile (right) of mitochondrial length in WT and *Prdm16*^{-/-} MEFs treated for 24 h with the Drp1 inhibitor, mDivi (30 μM), or vehicle. Bars, mean ± s.e.m.; *n* ≥ 16 fields from three biological replicates; **P* < 0.05 within length bins; one-way ANOVA with Bonferroni's post-hoc test within length bins. **d**, Fluorescence micrographs

showing fusion hybrids of WT or *Prdm16*^{-/-} MEFs transduced with baculovirus expressing mitochondria-tagged GFP and RFP before PEG-mediated fusion (scale bar 50 μm). Fused mitochondria are yellow (arrows), and were only observed in fusions of WT cells. **e**, western blot (upper, Supplementary Fig. 1 for full scans) and quantification of western blots for Mfn2 of WT and *Prdm16*^{-/-} MEFs. Bars, mean ± s.e.m.; *n* = 3 fields biological replicates; **P* < 0.05; two-tailed Student's *t*-test. **f**, Mitochondrial morphology in WT, *Prdm16*^{-/-} and *Mfn2*^{-/-} MEFs visualized by Tomm20 staining (red) (scale bar 20 μm). **(G)** Mitochondrial length in WT and *Prdm16*^{-/-} MEFs transduced with EV or Mfn2-IRES-GFP (Mfn2-GFP). Bars, mean ± s.e.m.; *n* ≥ 24 fields from three biological replicates; **P* < 0.05 compared to *Prdm16*^{-/-} EV in each length bin; one-way ANOVA with Bonferroni's post-hoc test within length bins.



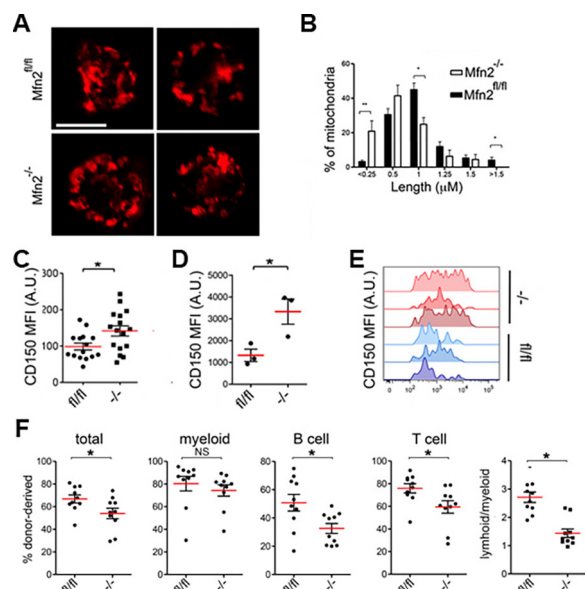
Extended Data Figure 3 | sPrdm16 interacts with the Mfn2 promoter. **a**, Schematic representation of Prdm16 protein domain structure. **b**, qPCR probe amplification scheme covering the Mfn2 promoter used in chromatin immunoprecipitation assays. **c**, Proximal Mfn2 promoter luciferase gene reporter assay in WT MEFs transfected 24h previously with flPrdm16, sPrdm16 or control vectors. Bars, mean \pm s.e.m.; $n = 3$ biological replicates; $*P < 0.05$; one-way ANOVA with Dunnett's post-hoc test. **d**, ChIP quantification in *Prdm16*^{-/-} MEFs transduced with retroviral Flag-Prdm16 constructs immunoprecipitated using Flag and TF2D antibodies. TFIID positive (GAPDH) and negative

(intergenic Chr 6) control probes are also shown. Quantification was performed after establishment of qPCR standard curves for all probes (see Methods). Bars, mean \pm s.e.m.; $n = 3$ technical replicates representative of two biological replicates; $*P < 0.05$; one-way ANOVA with Dunnett's post-hoc test. **e**, Percentage CD45.2 (donor) contribution in PB WBC in CD45.1⁺CD45.2⁺ mice reconstituted with 200 transduced (IRES-GFP or Mfn2-IRES-GFP) *Prdm16*^{+/-} CD45.2⁺ HSCs and 2×10^5 CD45.1⁺ competitor bone marrow cells. Plots, mean \pm s.e.m.; $n = 17$ recipients pooled from four independent transplants; $*P < 0.05$; two-tailed Student's *t*-test.

A**B**

Extended Data Figure 4 | Mitochondrial morphology of bone marrow populations. **a**, Representative images of mitochondrial morphology in Pham-reporter⁺ CD150^{hi} and CD150^{lo} HSCs (lin⁻Sca1⁺kit⁺Flt3⁻CD48⁻), MPPs (lin⁻Sca1⁺kit⁺CD48⁺), CMPs (lin⁻Sca1⁻kit⁺), CLPs (lin⁻Sca1^{lo}kit^{lo}IL7Rα⁺Flt3⁺) and lineage⁺ cells (scale bar, 5 μm).

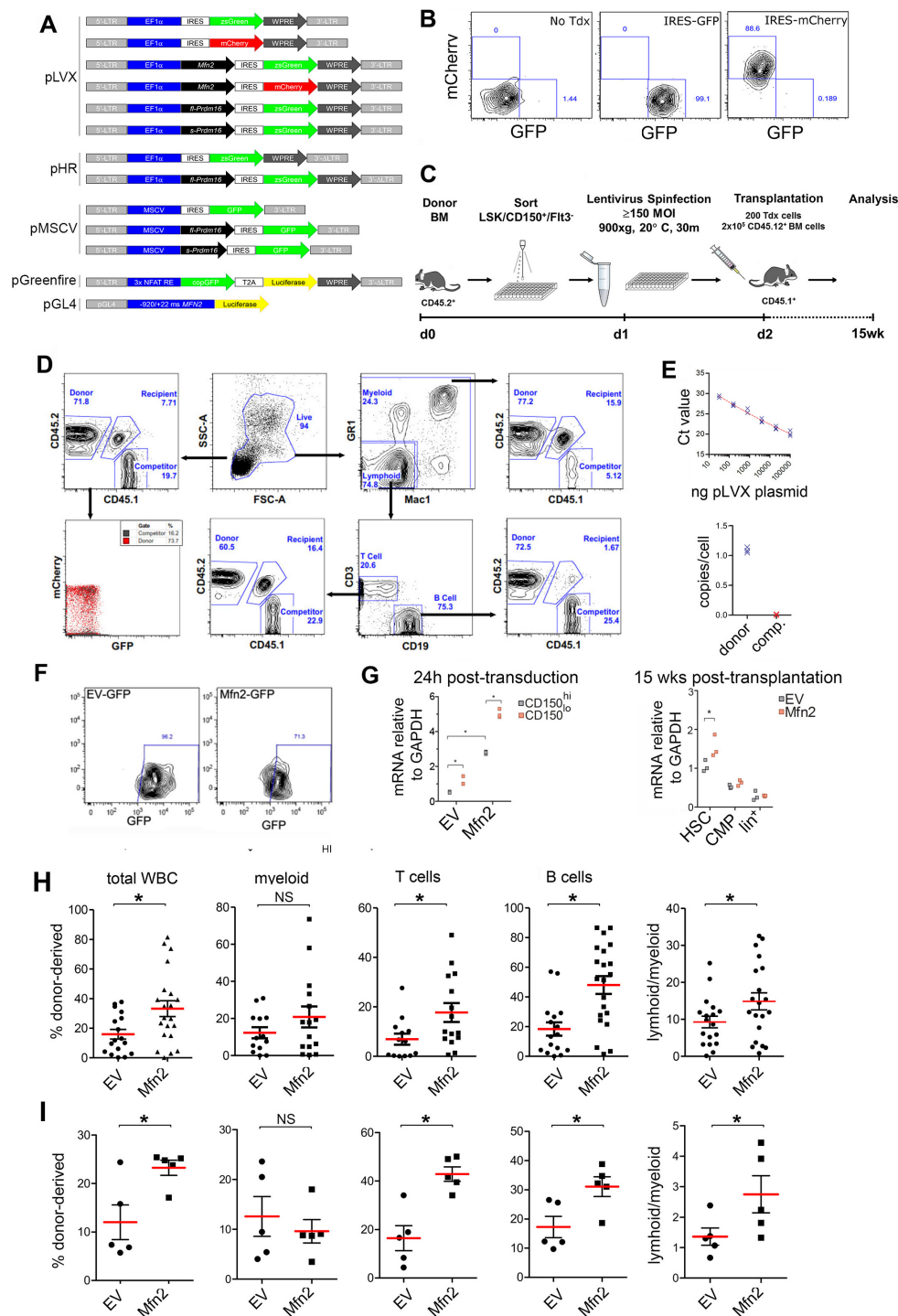
b, Mitochondrial length frequency distribution in Pham-reporter⁺ HSCs, progenitors and Lin⁺ cells. Bars, mean ± s.e.m.; $n \geq 14$ fields from three biological replicates; * $P < 0.05$ within length bins; one-way ANOVA with Bonferroni's post-hoc test within length bins.



Extended Data Figure 5 | Phenotype and function of *Mfn2^{-/-}* HSCs.

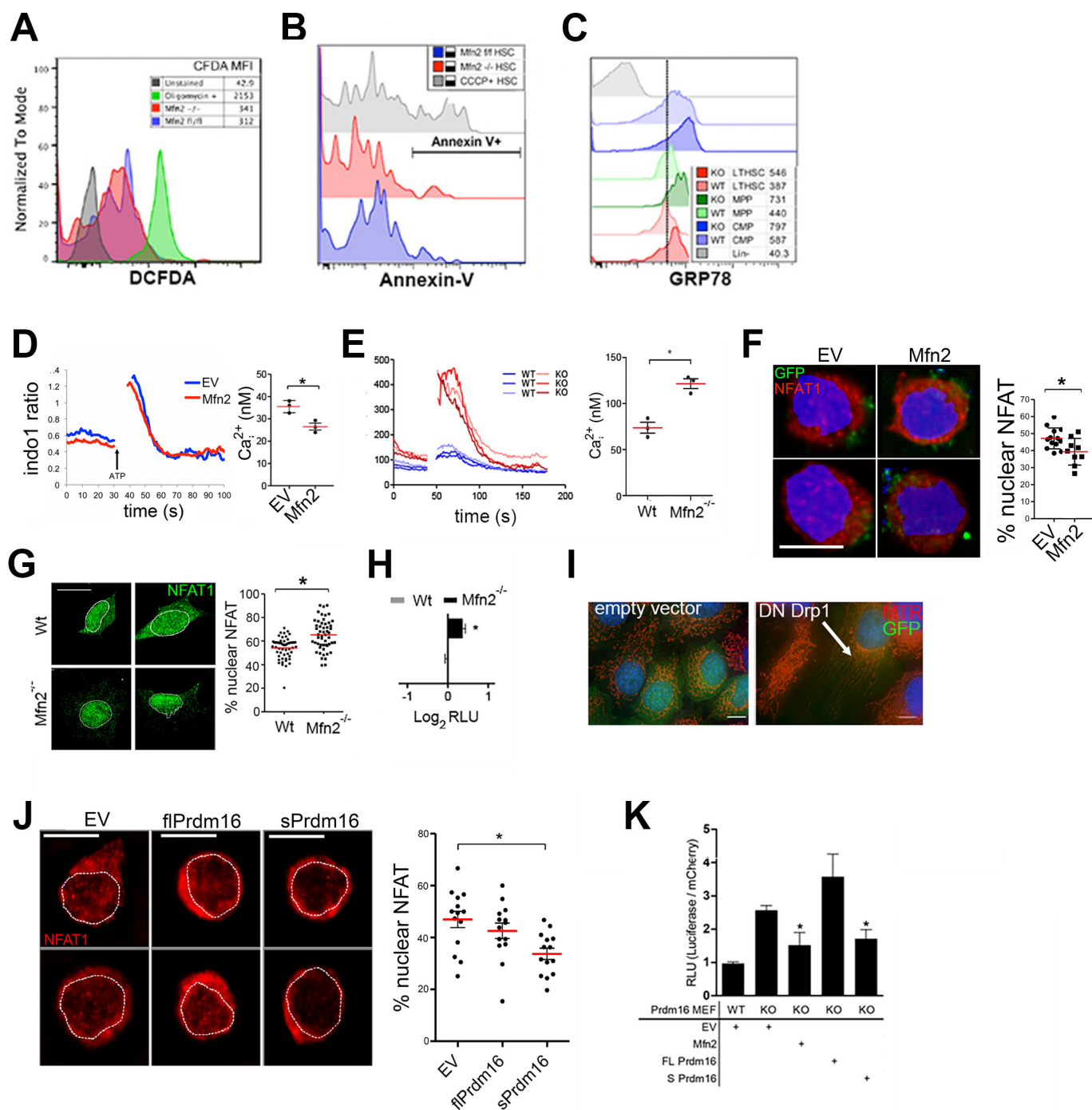
a, Mitochondrial morphology HSCs from *Mfn2^{fl/fl}* or *Mfn2^{fl/fl}-Vav-Cre* (*Mfn2^{-/-}*) mice visualized with immunostaining for Tomm20.

b, Frequency distribution of mitochondrial lengths (right panel, compared to WT in each length bin). Bars, mean \pm s.e.m.; $n \geq 7$ fields from two biological replicates; $*P < 0.05$ within length bins; two-tailed Student's *t*-test within length bins. **c**, CD150 surface staining mean fluorescence intensity (MFI) of HSCs (Lin⁻Sca1⁺kit⁺CD48⁻) in 8–12 week old *Mfn2^{fl/fl}* or *Mfn2^{fl/fl}-Vav-Cre* (*Mfn2^{-/-}*) mice. $n = 25$ biological replicates, $*P < 0.05$, two-tailed Student's *t*-test. **d**, **e**, CD150 MFI of *Mfn2^{fl/fl}* or *Mfn2^{fl/fl}-Vav-Cre* (*Mfn2^{-/-}*) donor HSCs 15 weeks after competitive transplantation. Plots in **d**, mean \pm s.e.m.; $n \geq 15$ mice from three biological replicates. **f**, $*P < 0.05$; two-tailed Student's *t*-test. **f**, Donor (CD45.2) chimerism 15 weeks after competitive transplantation of 2×10^5 *Mfn2^{fl/fl}* or *Mfn2^{fl/fl}-Vav-Cre* (*Mfn2^{-/-}*) fetal liver cells together with 2×10^5 CD45.1⁺ competitor bone marrow cells into CD45.1⁺CD45.2⁺ recipients. Plots, mean \pm s.e.m.; $n = 10$ mice pooled from two independent transplants; $*P < 0.05$; two-tailed Student's *t*-test.



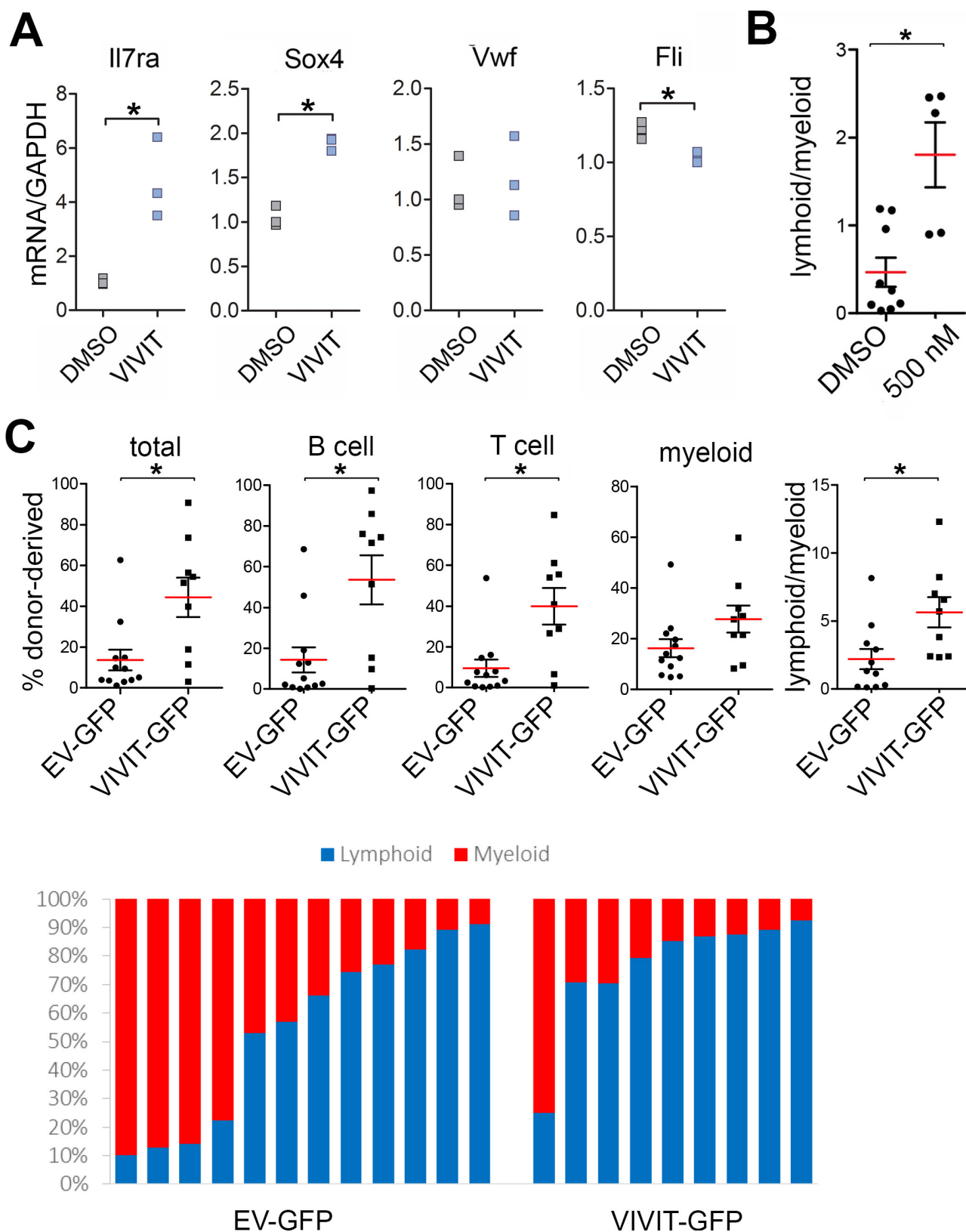
Extended Data Figure 6 | Analysis of competitive repopulation experiments and lentiviral vector transduction. **a**, Schematic representation of lentiviral constructs used. **b**, Representative example of transduction efficiency of purified HSCs with IRES-GFP and IRES-mCherry vectors 24–48 h post-transduction. **c**, Schematic representation of transduction and subsequent competitive repopulation experiments. **d**, Representative flow cytometric plots of the gates used to analyse recipients of competitive repopulation experiments. The CD45.1/CD45.2 gates were also applied to the myeloid, B and T cell gates to determine donor contribution to individual lineages. In this example, HSCs had been transduced with an IRES-GFP lentiviral vectors (efficiency 80%, see **b**). Although donor repopulation was high in the periphery (upper left panel, 71.8%) no GFP was detected (lower left panel), suggesting silencing of the vectors. **e**, Quantification of proviral copy number in donor-derived (CD45.2⁺) 15 weeks after transduction and competitive transplantation of 200 CD45.2⁺ HSCs transduced for 24 h with SIN LTR constructs pLVX-IRES-GFP or pLVX-Mfn2-IRES-GFP together with 2×10^5 CD45.1⁺ competitor fetal liver cells into lethally irradiated CD45.1⁺CD45.2⁺ recipients mice. Plots, mean \pm s.e.m.; $n \geq 19$ recipients from four transplantation experiments; * $P < 0.05$; two-tailed Student's *t*-test. **f**, Representative

example of transduction efficiency of purified HSCs with IRES-GFP and Mfn2-IRES-GFP lentiviral vectors. **g**, Expression of *Mfn2* mRNA in CD150^{hi} and CD150^{lo} HSCs 24 h post-transduction relative to EV control (left panel) and in donor-derived HSCs, CMPs and lineage⁺ cells 15 weeks after competitive transplantation of 200 transduced HSCs together with 2×10^5 CD45.1⁺ competitor cells (right panel). The data indicate partial silencing of the vector in HSCs, and complete silencing in the progenitors and more mature cells. $n = 3$ biological replicates; * $P < 0.05$; two-way analysis of variance (ANOVA) with Bonferroni's post-hoc test. **h**, Donor (CD45.2⁺) chimaerism 15 weeks after transplantation of 200 CD45.2⁺ HSCs transduced for 24 h with SIN LTR constructs pLVX-IRES-GFP or pLVX-Mfn2-IRES-GFP together with 2×10^5 CD45.1⁺ competitor fetal liver cells into lethally irradiated CD45.1⁺CD45.2⁺ recipients mice. Plots, mean \pm s.e.m.; $n \geq 19$ recipients from four transplantation experiments; * $P < 0.05$; two-tailed Student's *t*-test. **i**, Same experiments as in **h**, but using the non-SIN LTR vector, pHR. Plots, mean \pm s.e.m.; $n = 5$ recipients; * $P < 0.05$; two-tailed Student's *t*-test.



Extended Data Figure 7 | Nfat activity in *Mfn2*^{-/-} and *Prdm16*^{-/-} MEFs. **a**, *Mfn2*^{fl/fl} (blue) or *Mfn2*^{fl/fl}-Vav-Cre (*Mfn2*^{-/-}) (red) HSCs were stained with DCFDA to visualize intracellular reactive oxygen species levels. Cells were treated with oligomycin for 15 min as a positive control. **b**, HSC apoptosis levels were analysed by surface staining of Annexin V. **c**, Intracellular staining for the endoplasmic reticulum stress response chaperone GRP78 in *Mfn2*^{fl/fl} or *Mfn2*^{fl/fl}-Vav-Cre (*Mfn2*^{-/-}) HSCs (red colours) MPPs (green colours) and CMPs (blue colours). Histograms; representative of three biological replicates. **d**, Calcium flux trace (left) and baseline intracellular Ca^{2+} (right) in WT LSK cells transduced with IRES-GFP or *Mfn2*-IRES-GFP lentiviral vectors (**e**) and with WT or *Mfn2*^{-/-} MEFs stained with Indo1. Bars, mean \pm s.e.m.; $n = 3$ biological replicates; $*P < 0.05$; two-tailed Student's *t*-test. **f**, Nfat1 staining (left, scale bars, 5 μm) and fraction of nuclear Nfat1 (right, $*P < 0.05$) in HSCs transduced with IRES-GFP or *Mfn2*-IRES-GFP lentiviral vector. Bars, mean \pm s.e.m.; $n \geq 10$ from two biological replicates; $*P < 0.05$;

two-tailed Student's *t*-test. **g**, Nfat1 staining (left; scale bars, 20 μm) and fraction of nuclear Nfat1 (right; $*P < 0.05$) in WT or *Mfn2*^{-/-} MEFs. Bars, mean \pm s.d.; $n \geq 20$ fields from three independent experiments; $*P < 0.05$; two-tailed Student's *t*-test. **h**, Nfat luciferase gene reporter activity in WT or *Mfn2*^{-/-} MEFs. Bars, mean \pm s.d.; $n = 3$ experiments; $*P < 0.05$; two-tailed Student's *t*-test. **i**, Mitochondrial morphology after transfection of dominant negative Drp1 into HeLa cells (scale bars, 5 μm). Note extreme elongation of mitochondria (arrow), confirming that this vector is functional. **j**, Nfat1 staining (left, scale bars, 5 μm) and fraction of nuclear Nfat1 (right) in HSCs lentivirally transduced with sPrdm16 and flPrdm16. Plots, mean \pm s.e.m.; $n \geq 14$ fields of cells pooled from two biological replicates; $*P < 0.05$; one-way ANOVA with Dunnett's post-hoc test. **k**, Nfat luciferase gene reporter activity in WT and in *Prdm16*^{-/-} MEFs transfected with *Mfn2*, flPrdm16 or sPrdm16 constructs. Bars, mean \pm s.d.; $n = 3$ experiments; $*P < 0.05$; one-way ANOVA with Dunnett's post-hoc test.



Extended Data Figure 8 | Effect of VIVIT on WT HSCs. **a**, Expression of lymphoid (*Sox4* and *Il7r*) and myeloid/platelet (*Vwf* and *Fli1*) genes in HSCs treated with VIVIT or DMSO for 24 h. $n = 3$ biological replicates; $*P < 0.05$; Student's *t*-test. **b**, Lymphoid/myeloid ratio of CD45.2⁺ HSCs cultured for 4 days in DMSO and 500 nM VIVIT, transplanted with 2×10^5 CD45.1⁺CD45.2⁺ competitor bone marrow cells into CD45.1⁺ recipients and analysed 15 weeks post-transplant for lymphoid/myeloid ratio of donor compartment. Plots, mean \pm s.e.m.; $n = 5$ –9 recipients

from two transplant experiments; $*P < 0.05$; two-tailed Student's *t*-test. **c**, Donor chimaerism analysis 15 weeks after transplantation of CD45.2⁺ HSCs transduced with IRES-GFP or VIVIT-GFP with 2×10^5 CD45.1⁺CD45.2⁺ competitor bone marrow cells into CD45.1⁺ recipients. Plots, mean \pm s.e.m.; $n = 9$ –12 recipients from two independent transplant experiments; $*P < 0.05$; two-tailed Student's *t*-test. Bottom panel shows lymphoid and myeloid donor contribution in each individual recipient.

Extended Data Table 1 | Phenotypic analysis of the haematopoietic system in *Mfn2^{fl/fl}*-Vav-Cre mice

</

a–f, Analysis of 8–10 week-old *Mfn2^{fl/fl}* or *Mfn2^{fl/fl}*-Vav-Cre mice for (a) frequency of LSK cells and HSCs, (b) peripheral blood counts, (c) bone marrow progenitor populations (MPPs (lin[−]Sca1⁺kit⁺CD48⁺), CMPs (lin[−]Sca1⁺kit⁺CD16/32^{low}CD34[−]), CLPs (lin[−]Sca1⁺kit⁺IL7Ra⁺Flt3⁺), MEPs (Lin[−]Kit⁺Sca1[−]CD16/32^{low}CD34[−])), (d) myeloid (Gr1⁺/Mac1⁺), T (CD3⁺) and B (CD19⁺) cells in the spleen, (e) myeloid (Gr1⁺/Mac1⁺), T (CD3⁺) and B (CD19⁺) cells in peripheral blood mononuclear cells, and (f) thymocyte subpopulations (DP: CD4⁺CD8⁺, DN: CD4⁺CD8[−], DN1: CD4⁺CD8[−]CD44⁺CD25[−], DN2: CD4⁺CD8[−]CD44⁺CD25⁺, DN3: CD4⁺CD8[−]CD44[−]CD25⁺, DN4: CD4⁺CD8[−]CD44[−]CD25[−]). Data are mean ± s.e.m.; n = 3 mice; *P < 0.05; two-tailed Student's t-test.

Extended Data Table 2 | Statistical analysis of limiting dilution experiments shown in Fig. 3c

total reconstitution				lymphoid reconstitution				myeloid reconstitution			
Group	Lower	Estimate	Upper	Group	Lower	Estimate	Upper	Group	Lower	Estimate	Upper
KO	108.1	55.6	28.72	KO	407.6	153.2	57.8	KO	35.8	20.23	11.5
WT	30.3	17	9.59	WT	69.8	38.2	21	WT	18.5	8.96	4.5
Chisq	DF	P.value		Chisq	DF	P.value		Chisq	DF	P.value	
7.64	1	0.00571		6.71	1	0.00957		3.1	1	0.0785	

Graded Foxo1 activity in T_{reg} cells differentiates tumour immunity from spontaneous autoimmunity

Chong T. Luo^{1,2}, Will Liao³, Saida Dadi¹, Ahmed Toure¹ & Ming O. Li¹

Regulatory T (T_{reg}) cells expressing the transcription factor Foxp3 have a pivotal role in maintaining immunological self-tolerance^{1–5}; yet, excessive T_{reg} cell activities suppress anti-tumour immune responses^{6–8}. Compared to the resting T_{reg} (rT_{reg}) cell phenotype in secondary lymphoid organs, T_{reg} cells in non-lymphoid tissues exhibit an activated T_{reg} (aT_{reg}) cell phenotype^{9–11}. However, the function of aT_{reg} cells and whether their generation can be manipulated are largely unexplored. Here we show that the transcription factor Foxo1, previously demonstrated to promote T_{reg} cell suppression of lymphoproliferative diseases^{12,13}, has an unexpected function in inhibiting aT_{reg}-cell-mediated immune tolerance in mice. We find that aT_{reg} cells turned over at a slower rate than rT_{reg} cells, but were not locally maintained in tissues. aT_{reg} cell differentiation was associated with repression of Foxo1-dependent gene transcription, concomitant with reduced Foxo1 expression, cytoplasmic localization and enhanced phosphorylation at the Akt sites. T_{reg}-cell-specific expression of an Akt-insensitive Foxo1 mutant prevented downregulation of lymphoid organ homing molecules, and impeded T_{reg} cell homing to non-lymphoid organs, causing CD8⁺ T-cell-mediated autoimmune diseases. Compared to T_{reg} cells from healthy tissues, tumour-infiltrating T_{reg} cells downregulated Foxo1 target genes more substantially. Expression of the Foxo1 mutant at a lower dose was sufficient to deplete tumour-associated T_{reg} cells, activate effector CD8⁺ T cells, and inhibit tumour growth without inflicting autoimmunity. Thus, Foxo1 inactivation is essential for the migration of aT_{reg} cells that have a crucial function in suppressing CD8⁺ T-cell responses; and the Foxo signalling pathway in T_{reg} cells can be titrated to break tumour immune tolerance preferentially.

rT_{reg} cells, defined by high expression of CD62L and low expression of CD44, were abundant in lymph nodes and spleens, whereas CD62L^{lo}CD44^{hi} aT_{reg} cells were present in both lymphoid organs and non-lymphoid tissues such as the liver and lamina propria of the intestine (Extended Data Fig. 1a). To examine how T_{reg} cells are maintained in these tissues, we connected congenitally marked C57BL/6 mice using parabiosis (Extended Data Fig. 1b). In line with a recent study¹⁴, rT_{reg} cells as well as naive CD4⁺ T cells reached chimaerism of approximately 50%, and aT_{reg} cells, in particular lamina propria T_{reg} cells, were skewed towards the host 2 weeks after surgery (Fig. 1a). Nevertheless, in contrast to liver-resident CD49a⁺NK1.1⁺ cells, all T_{reg} cell populations were mixed by 4 weeks (Fig. 1a), revealing that they were not locally sustained for an extended period.

Antigen-experienced conventional T cells that recirculate through blood, lymph and non-lymphoid tissues can be short-lived effector cells or long-lived effector memory cells¹⁵. To dissect the homeostatic properties of T_{reg} cells, we disconnected the parabionts after 4 weeks, and assessed the turnover of cells originated from the non-host parabiont (Extended Data Fig. 1b, c). Lymph node or splenic rT_{reg} cells turned over at a rate close to that of naive CD4⁺ T cells, with a decay half time of between 4 and 7 weeks (Fig. 1b). In contrast, aT_{reg} cells from these

tissues turned over at a substantially slower rate, with a half time of around 14 weeks (Fig. 1b). Notably, liver or lamina propria T_{reg} cells had a comparable decay rate of between 13 and 15 weeks (Fig. 1b). Thus, compared to rT_{reg} cells, aT_{reg} cells from both lymphoid and non-lymphoid tissues turn over more slowly, resembling effector memory T cells.

We wanted to determine how aT_{reg} cell trafficking and homeostasis are regulated, and whether these processes can be manipulated. The transcription factor Foxo1 integrates diverse environmental signals to control T-cell homeostasis and differentiation^{16,17}. Expression of Foxo1 is essential for T_{reg} cell function^{12,18}, but its role in aT_{reg} and rT_{reg} cell subsets has not been defined. To this end, we performed gene-expression profiling experiments of splenic aT_{reg} and rT_{reg} subsets. By cross-referencing the differentially expressed genes and

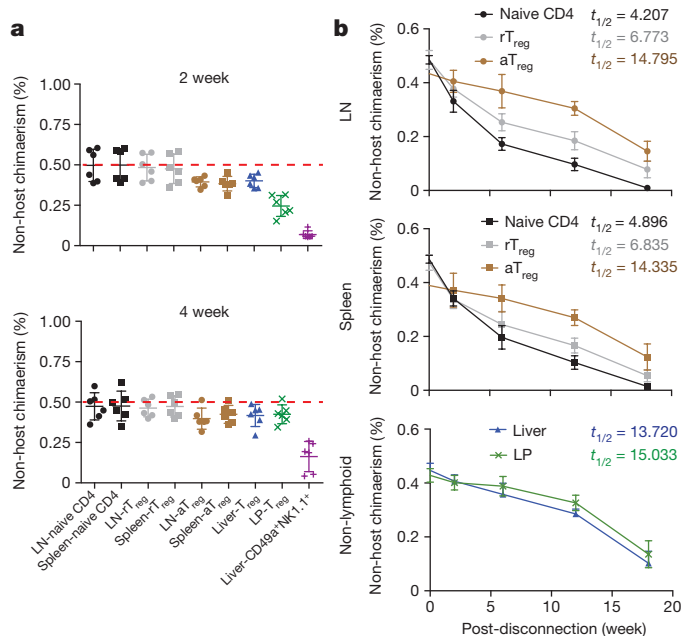


Figure 1 | aT_{reg} cells have a slow turnover, but are not locally maintained in non-lymphoid tissues. **a**, The frequencies of non-host derived cells in parabiotic mice 2 or 4 weeks after surgery, including naive CD4⁺ T cells (CD4⁺Foxp3⁻CD62L^{hi}CD44^{lo}), rT_{reg} cells (CD4⁺Foxp3⁺CD62L^{hi}CD44^{lo}) and aT_{reg} cells (CD4⁺Foxp3⁺CD62L^{lo}CD44^{hi}) in the lymph node (LN) and spleen, total T_{reg} cells (CD4⁺Foxp3⁺) in the liver and colon lamina propria (LP), and CD49a⁺NK1.1⁺ innate lymphoid cells in the liver. *n* = 6. **b**, Parabionts were separated 4 weeks after connection, and the percentages of non-host chimaerism at 2, 6, 12 and 18 weeks after separation are shown. *t*_{1/2} depicts the amount of time it took the population to decay to half of its original size. *n* = 4 for each time point. Data are mean ± s.e.m. Comparison between aT_{reg} and rT_{reg}: *P* = 0.0025 for lymph node; *P* = 0.0335 for spleen (two-way analysis of variance (ANOVA)).

¹Immunology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ²Louis V. Gerstner Jr Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ³New York Genome Center, New York, New York 10013, USA.

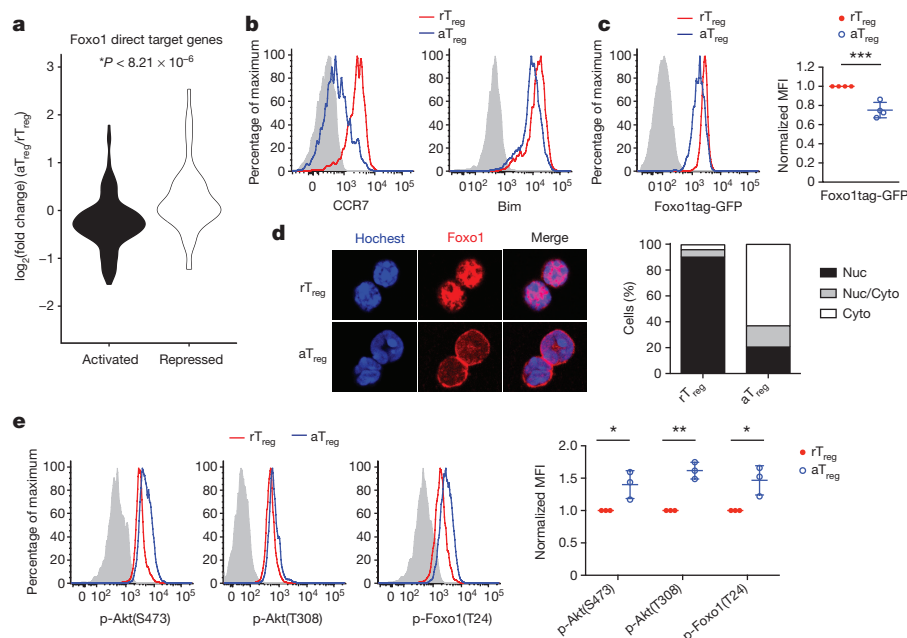


Figure 2 | aT_{reg} cell differentiation is associated with downregulation of Foxo1-dependent gene expression. **a**, Gene expression comparison of Foxo1 direct target genes in splenic aT_{reg} (CD62L^{lo}CD44^{hi}) versus rT_{reg} (CD62L^{hi}CD44^{lo}) cells. Foxo1 direct target genes are defined as: (1) being differentially expressed between wild-type and Foxo1 knockout T_{reg} cells; (2) having their expression corrected by Foxo1CA; and (3) having Foxo1 recruited to the gene locus. **b**, **c**, **e**, Flow cytometric analysis of Foxo1-activated target genes CCR7 and Bim (**b**), Foxo1 protein (using

green fluorescent protein (GFP) as a reporter; Foxo1tag-GFP (**c**), and phosphorylated Akt and Foxo1 (p-Akt, and p-Foxo1) (**e**) in splenic aT_{reg} and rT_{reg} cells. Mean \pm s.e.m. of mean fluorescence intensity (MFI), normalized to rT_{reg} cells, are shown. $n = 4$ (**c**); $n = 3$ (**e**). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$ (paired *t*-test). **d**, Immunofluorescence staining of Foxo1 in splenic rT_{reg} and aT_{reg} cells. $n = 70$. Cyto, Nuc and Nuc/Cyto denote Foxo1 expression in the cytoplasm, nucleus and both the nucleus and cytoplasm, respectively. Original magnification, $\times 60$.

the Foxo1-regulated genes¹², we found that aT_{reg} or rT_{reg} cells preferentially expressed the Foxo1-downregulated or -upregulated transcripts, respectively (Extended Data Fig. 2a and Supplementary Table). Furthermore, in reference to a Foxo1 direct target gene signature¹², the Foxo1-repressed or -activated transcripts were enriched in aT_{reg} or rT_{reg} cells, respectively (Fig. 2a and Supplementary Table). Notably, several Foxo1-activated genes that promote lymphoid organ homing, including *Klf2*, *Ccr7* and *S1pr1*, were highly expressed in rT_{reg} cells, whereas the Foxo1-repressed genes potentially involved in T-cell migration or retention in tissues, such as *Lamc1*, *Nid2* and *Mmp9*, were induced in aT_{reg} cells (Extended Data Fig. 2b, c and Supplementary Table). In line with RNA-sequencing data, protein levels of two well-defined Foxo1-activated targets, *Ccr7* and *Bim*, were downregulated in aT_{reg} cells (Fig. 2b), supporting that Foxo1-dependent gene expression is repressed in aT_{reg} cells.

Compared to rT_{reg} cells, aT_{reg} cells expressed lower amounts of Foxo1 transcript and protein (Fig. 2c and Supplementary Table). Foxo1 nuclear localization and protein stability is attenuated after phosphorylation by Akt^{16,17}. Indeed, while Foxo1 predominately resided in the nucleus of rT_{reg} cells, it translocated to the cytoplasm of aT_{reg} cells (Fig. 2d). Phosphorylation of Akt, Foxo1 and the mTORC1 signalling pathway marker S6 ribosomal protein was increased in aT_{reg} cells (Fig. 2e and Extended Data Fig. 2d). These findings demonstrate that aT_{reg} cell differentiation is associated with activation of the Akt kinase, with concomitant repression of Foxo1 nuclear localization and expression.

Inclusion of an Akt inhibitor prevented the induction of aT_{reg} from rT_{reg} cells *in vitro* (Extended Data Fig. 2e). To determine the specific role of Foxo1 inactivation in aT_{reg} cells *in vivo*, we used a mouse strain expressing a mutant form of Foxo1 that is refractory to Akt-triggered inhibition¹². Mice carrying the Rosa26-floxed stop mutant *Foxo1* allele were bred to the *Foxp3^{Cre}* background to induce T_{reg}-cell-specific expression of a constitutively active form of Foxo1 (ref. 12), herein designated CA. CA was expressed at increasing levels in

CA/+ (*Foxp3^{Cre}Foxo1CA/+*) or CA/CA (*Foxp3^{Cre}Foxo1CA/Foxo1CA*) T_{reg} cells, and was constitutively localized in the nucleus (Extended Data Fig. 3a, b). As expected, CA triggered a dose-dependent increase of its target gene expression (Extended Data Fig. 3c). Thymic T_{reg} cell differentiation was unperturbed in CA/+ or CA/CA mice (data not shown). However, CD62L^{lo} aT_{reg} phenotype cells in lymph nodes were proportionally decreased in 9–12-day-old CA/+ or CA/CA mice (Fig. 3a), in line with a role for Foxo1 in inducing CD62L expression^{19–21}. T-cell activation markers CD69 and ICOS were comparably induced among T_{reg} cells from wild-type, CA/+ or CA/CA mice, with the exception that a higher fraction of CD62L^{hi} rT_{reg} phenotype cells expressed these molecules in CA/+ or CA/CA mice (Fig. 3b and Extended Data Fig. 3d). Similarly, CD62L^{hi} rT_{reg} phenotype cells in CA/CA mice expressed higher levels of the activation-associated genes, including *Cd69*, *Egr2* and *Il1r2*, than wild-type rT_{reg} cells (Extended Data Fig. 3e). In addition, EdU labelling experiments showed that CA/CA rT_{reg} phenotype cells had a higher proliferation rate than wild-type rT_{reg} cells (Extended Data Fig. 3f).

To dissect the effect of CA on T_{reg} cells further, we purified CD62L^{hi}CD44^{lo}CD69⁺ICOS⁺ rT_{reg} cells from wild-type, CA/+ and CA/CA mice, and performed *in vitro* culture experiments. Compared to wild-type rT_{reg} cells, the CA-expressing rT_{reg} cells showed no defect in activation, proliferation or survival, but they failed to downregulate CD62L (Extended Data Fig. 4). CCR7 downregulation was also attenuated in CA-expressing T_{reg} cells (Extended Data Fig. 3c), revealing that relief of Akt-triggered Foxo1 inhibition is sufficient to maintain high expression of molecules involved in lymph node homing and intranodal T-cell migration. To determine whether such alteration perturbed T_{reg} cell trafficking, we intravenously administered a CD4 antibody, which predominantly labelled cells in the red pulp or marginal zone of spleen, but not the white pulp. Fewer CA-expressing T_{reg} cells were found among the labelled red pulp/marginal zone fraction (Fig. 3c), whereas comparable percentages of conventional T cells were labelled (Extended Data Fig. 5a), suggesting that CA

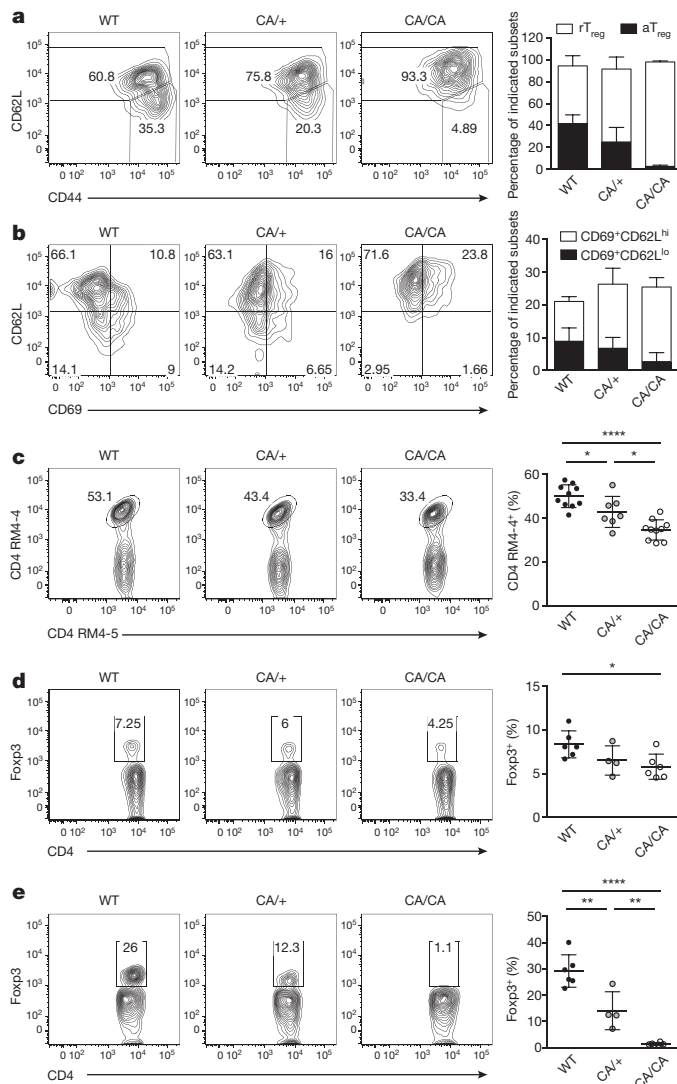


Figure 3 | Expression of a constitutively active form of Foxo1 in T_{reg} cells alters cell migration. **a, b**, Flow cytometric analysis of CD44, CD62L (**a**), and CD69, CD62L (**b**) in lymph node T_{reg} cells (CD4⁺Foxp3⁺) from wild-type (WT), *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice. Quantification of rT_{reg} (CD62L^{hi}CD44^{lo}) and aT_{reg} (CD62L^{lo}CD44^{hi}) cell subsets (**a**), and CD69⁺CD62L^{hi}, CD69⁺CD62L^{lo} populations (**b**) are shown, with percentages of indicated populations among total T_{reg} cells. *n* = 6. **c**, *In vivo* labelling experiment: anti-CD4-biotin (RM4-4) was injected intravenously 5 min before analysis. Anti-CD4 (RM4-5) was used to label total CD4⁺ T cells. Representative plots from gated T_{reg} cells (CD4 RM4-5⁺Foxp3⁺) are shown. Quantification shows percentage of RM4-4⁺ cells among T_{reg} cells. *n* = 10 (WT, CA/CA); *n* = 7 (CA/+). **d, e**, Foxp3 expression in lymph node (**d**) or liver (**e**) CD4⁺ T cells. Quantification shows percentages of T_{reg} cells among total CD4⁺ T cells. *n* = 6 (WT, CA/CA); *n* = 4 (CA/+). All mice were 9–12 days old. Data are mean ± s.e.m. Results represent at least three independent experiments. **P* < 0.05; ***P* < 0.01; *****P* < 0.0001 (unpaired *t*-test).

prevented the migration of T_{reg} cells from white pulp to red pulp/marginal zone.

rT_{reg} and aT_{reg} cells reside at different locations in secondary lymphoid organs, and engage discrete mechanisms of homeostatic maintenance¹⁴. Whereas the total proportion or number of lymph node CD62L^{hi} T_{reg} cells was similar or increased in CA-expressing mice compared to wild-type mice, CD62L^{lo} T_{reg} cells were reduced with increasing doses of CA expression (Extended Data Fig. 5b, c), resulting in reduced total lymph node T_{reg} cells (Fig. 3d). Thus, although

Foxo1 inactivation is not essential for T_{reg} cell activation, proliferation or survival, it is required to control the expression of trafficking molecules that may promote aT_{reg} cells to migrate away from the rT_{reg} cell niche and further expand. Indeed, CA expression from one or two alleles caused a dose-dependent reduction of liver T_{reg} cells (Fig. 3e).

To examine whether the aT_{reg} cell defects were sustained beyond the neonatal stage, we analysed 4–6-week-old mice. Expression of CA triggered a dose-dependent reduction of aT_{reg} cells in lymphoid organs as well as in non-lymphoid tissues, leaving rT_{reg} cells less affected (Fig. 4a and Extended Data Fig. 6a, b). Compared to wild-type and CA/+ mice, all CA/CA mice succumbed to a wasting disease and death by 4 months of age (Fig. 4b and Extended Data Fig. 6c, d). In contrast to T_{reg}-cell-specific Foxo1-deficient mice¹², CA/CA mice did not exhibit the typical inflammatory phenotype associated with the ‘Scurfy’ mutation of *Foxp3* such as tail crusting, splenomegaly or lymphadenopathy (Fig. 4c and Extended Data Fig. 6c, e). Expression of CA did not perturb the levels of the T_{reg} cell markers CTLA4, LAG3 and GITR, or T_{reg} cell suppressive activity *in vitro* (Extended Data Fig. 6f, g). Yet, a dense infiltrate of leukocytes were observed in many organs of CA/CA mice, including the liver and colon (Fig. 4d). Serum alanine aminotransferase (ALT) activity, a biomarker for liver injury, as well as colon pathology scores were increased (Fig. 4e), revealing that the immunopathology had resulted in tissue damage.

The autoimmune lethal phenotype in CA/CA mice was associated with activation of CD4⁺ and CD8⁺ T cells (Extended Data Fig. 7a, d), yet they produced only modestly higher or comparable amounts of inflammatory cytokines compared to T cells from wild-type or CA/+ mice (Extended Data Fig. 7b, c, e, f). Splenic CD8⁺ T cells, however, expressed substantially higher levels of the cytolytic molecule granzyme B (Gzmb) (Fig. 4f, g). Much enhanced expression of Gzmb, but not IFN- γ , was also observed in liver- and lamina-propria-infiltrating CD8⁺ T cells (Fig. 4f, g and Extended Data Fig. 8a, b). In addition, T-cell populations were skewed towards CD8⁺ T cells in these tissues (Extended Data Fig. 8c), suggesting that augmented effector CD8⁺ T-cell responses might trigger the immunopathology. To test this hypothesis, we crossed CA/CA mice to the CD8-deficient background. Depletion of CD8⁺ T cells completely rescued the wasting disease and lethal phenotype of CA/CA mice (Fig. 4h and Extended Data Fig. 9a). Serum ALT levels and tissue pathology were also fully rectified (Fig. 4i and Extended Data Fig. 9b). Collectively, these findings suggest that aT_{reg} cells have a particularly important function in suppressing CD8⁺ T-cell-mediated tissue destruction.

Spontaneous autoimmune pathology was observed in CA/CA but not CA/+ mice, in line with CA dose-dependent repression of aT_{reg} cells (Fig. 4a and Extended Data Fig. 6a, b). CA expression also led to varying degrees of T_{reg} cell reduction in different tissues, with the lamina propria more affected than the liver (Fig. 4a). In fact, the magnitude of T_{reg} cell loss in lymphoid organs and non-lymphoid tissues was inversely correlated with Foxo1 activity, as revealed by expression of two direct Foxo1 target genes (Fig. 5a).

High numbers of T_{reg} cells are found in tumours^{6–8}, which is associated with poor prognosis of cancer patients²². The role of T_{reg} cells in tumour immunity has been investigated primarily with methods that deplete T_{reg} cells in all tissues^{23,24}. To study how CA affects tumour-infiltrating T_{reg} cells, we used the MMTV-PyMT (PyMT) spontaneous mammary tumour model²⁵. Similar to T_{reg} cells from the liver and lamina propria, tumour-infiltrating T_{reg} cells exhibited an activated phenotype (Extended Data Fig. 10a), yet they expressed the lowest level of Foxo1 targets (Fig. 5a), suggesting that they might be most sensitive to Foxo1 gain-of-function. To test this hypothesis, we crossed CA/+ mice to the PyMT background. Indeed, tumour-infiltrating T_{reg} cells were diminished by 2.6-fold with CA expression from one allele (Fig. 5b, c), which resulted in profound inhibition of tumour growth (Fig. 5d). Although CD8⁺ T cells were phenotypically indistinguishable in healthy tissues from CA/+ and wild-type mice (Fig. 4f, g),

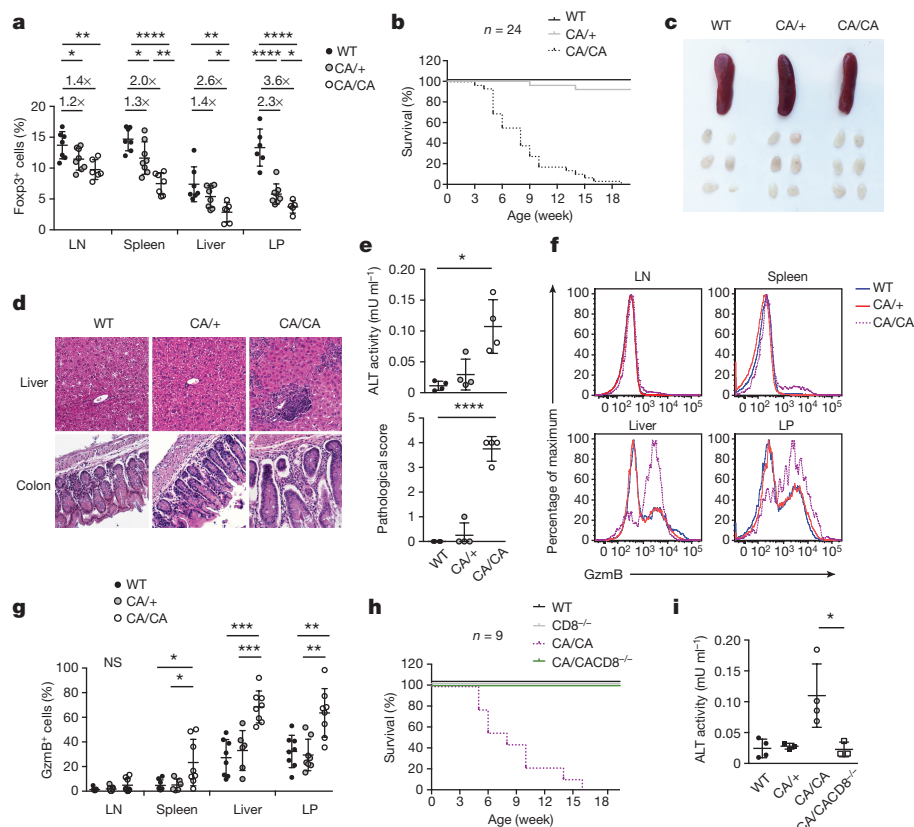


Figure 4 | Foxo1 hyperactivation in T_{reg} cells causes a CD8⁺ T-cell-dependent inflammatory disease. **a**, The frequencies of T_{reg} cells among CD4⁺ T cells in lymph node, spleen, liver and colon lamina propria of wild-type, *Foxp3*^{Cre}*Foxo1*CA/+ (CA/+) or *Foxp3*^{Cre}*Foxo1*CA/*Foxo1*CA (CA/CA) mice. $n = 6$ (WT, CA/CA); $n = 8$ (CA/+). Numbers above plots indicate fold changes in comparison to wild type. **b**, Survival of wild-type, CA/+ and CA/CA mice. **c**, Representative image of spleens and peripheral (axillary, brachial and inguinal) lymph nodes. **d**, Haematoxylin and eosin staining of liver and colon sections. Original magnification, $\times 20$. **e**, Top, serum ALT activity; bottom, histological grading of colons. $n = 4$. **f**, Flow cytometric analysis of granzyme B (GzmB) expression in CD8⁺ cells (f) and quantification (g) from lymph node, spleen, liver and lamina propria of wild-type, CA/+ and CA/CA mice. In **g**, $n = 8$ (WT, CA/CA); $n = 6$ (CA/+). NS, not significant. **h**, Survival of wild-type, CD8^{-/-} (CD8^{-/-}; also known as *Cd8a*^{-/-}), CA/CA and *Foxp3*^{Cre}*Foxo1*CA/*Foxo1*CA *CD8*^{-/-} (CA/CA^{CD8}^{-/-}) mice. **i**, Serum ALT activity from wild-type, CD8^{-/-}, CA/CA and CA/CA^{CD8}^{-/-} mice. $n = 4$. All mice were 4–6 weeks old. Data are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$ (unpaired *t*-test).

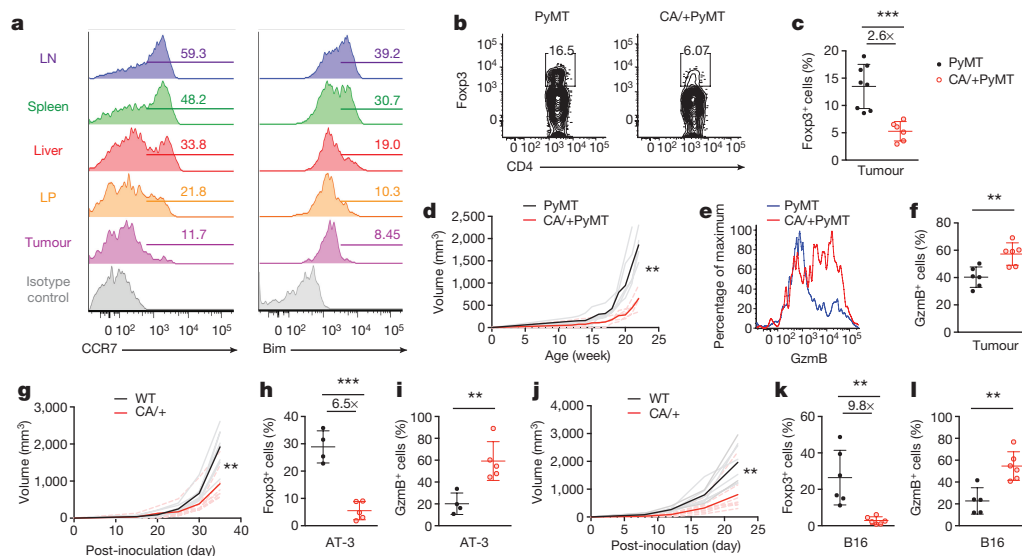


Figure 5 | Tuned activation of Foxo1 in T_{reg} cells results in enhanced anti-tumour immunity without inflicting autoimmunity. **a**, Flow cytometric analysis of Foxp3⁺ cells in lymph node, spleen, liver, colon lamina propria and tumour of 22–24-week-old PyMT mammary tumour-bearing mice. Numbers indicate percentage of CCR7- or Bim-positive cells shown in the gate. **b**, Foxp3 expression in tumour-infiltrating CD4⁺ T cells from PyMT or *Foxp3*^{Cre}*Foxo1*CA/+PyMT (CA/+PyMT) mice. **c**, The frequencies of T_{reg} cells among CD4⁺ T cells in tumours of PyMT or CA/+PyMT mice. $n = 8$ (PyMT); $n = 6$ (CA/+PyMT). Number above plots indicates fold change in comparison to PyMT. **d**, Tumour growth curve of PyMT and CA/+PyMT mice. $n = 5$. **e**, Flow cytometric analysis of GzmB expression in CD8⁺ cells (e) and quantification (f) from tumours of PyMT and CA/+PyMT

mice. $n = 6$. **g**–**i**, Eight-to-ten-week-old wild-type or *Foxp3*^{Cre}*Foxo1*CA/+ (CA/+) mice received orthotopic inoculation of PyMT-derived mammary tumour cells (AT-3) (g–i), or subcutaneous injection of B16 melanoma cells (j–l). **g**, j, Tumour growth curve of wild-type and CA/+ mice. $n = 8$. **h**, k, The frequencies of T_{reg} cells among CD4⁺ T cells in tumours of wild-type and CA/+ mice. In **h**, $n = 4$ (WT); $n = 5$ (CA/+). In **k**, $n = 6$. Numbers above plots indicate fold change in comparison to wild type. **i**, l, The frequencies of GzmB-expressing CD8⁺ T cells from tumours of wild-type and CA/+ mice. In **i**, $n = 4$ (WT); $n = 5$ (CA/+). In **l**, $n = 5$ (WT); $n = 6$ (CA/+). Results represent at least three independent experiments. Data are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$ (unpaired *t*-test (c, f, h, i, k, l) and two-way analysis of variance (ANOVA) (d, g, j)).

tumour-infiltrating CD8⁺ T cells from CA/+PyMT mice expressed higher amounts of GzmB (Fig. 5e, f).

Tumour progression in PyMT mice is accompanied by loss of genome stability²⁶, which induces accumulation of mutations to foster cell transformation. To determine whether CA expression in T_{reg} cells could inhibit growth of already transformed cells, we used syngeneic AT-3 cells derived from PyMT mice. Orthotopic implantation of AT-3 to the mammary fat pad resulted in aggressive tumour growth and lethality in wild-type mice, both of which were attenuated in CA/+ mice (Fig. 5g and Extended Data Fig. 10b). CA expression led to a 6.5-fold reduction of tumour-infiltrating T_{reg} cells (Fig. 5h and Extended Data Fig. 10c), which was associated with increased expression of GzmB in tumour-infiltrating CD8⁺ T cells (Fig. 5i and Extended Data Fig. 10d). Importantly, depletion of CD8⁺ T cells by crossing CA/+ mice to the CD8-deficient background restored tumour growth (Extended Data Fig. 10e), revealing an essential function for CD8⁺ T cells in tumour suppression. To investigate whether the CA effect was applicable to tumours of other tissue origin, we used B16 melanoma cells. Similar to AT-3, B16 growth was inhibited in CA/+ mice (Fig. 5j), which was accompanied by an approximately tenfold reduction of tumour-infiltrating T_{reg} cells and increased GzmB-expressing CD8⁺ T cells (Fig. 5k, l). These findings reveal that tumour-associated T_{reg} cells are generally more susceptible to CA-triggered depletion.

rT_{reg} and aT_{reg} cell subsets are well defined in humans and mice^{9,10}, but their individual contributions to immune tolerance have been enigmatic. Our data reveal that Foxo1 repression is a major transcriptional reprogramming event associated with differentiation of aT_{reg} cells, which supports their distinct migration pattern resembling that of effector memory T cells. Using a Foxo1 gain-of-function model that preferentially triggers aT_{reg} loss in peripheral tissues, we could identify a prominent role for these cells in the control of CD8⁺ T-cell tolerance. Although CD4⁺ T cells are insufficient to induce immunopathology, aT_{reg} cells might inhibit their helper function to suppress cytotoxic T-cell responses. In addition, aT_{reg} cells may inhibit CD8⁺ effector T cells by modulating antigen-presenting cells as shown in a tumour model²⁷. Importantly, tumour-associated aT_{reg} cells are more susceptible to Foxo1-triggered depletion as a likely consequence of higher antigen load, and effective tumour immunity can be induced without overt spontaneous autoimmunity. aT_{reg} cell differentiation is dependent on T-cell receptor signalling^{28,29} that inactivates Foxo1 via Akt. Drugs that target upstream TCR signalling molecules, as recently shown for phosphatidylinositol-3-OH kinase (PI(3)K)³⁰, may impinge on the Foxo1 pathway to break T_{reg}-cell-mediated tumour immune tolerance selectively.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 July; accepted 23 November 2015.

Published online 20 January 2016.

1. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T cells and immune tolerance. *Cell* **133**, 775–787 (2008).
2. Shevach, E. M. Biological functions of regulatory T cells. *Adv. Immunol.* **112**, 137–176 (2011).
3. Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30**, 531–564 (2012).
4. Benoist, C. & Mathis, D. Treg cells, life history, and diversity. *Cold Spring Harb. Perspect. Biol.* **4**, a007021 (2012).
5. Bluestone, J. A., Bour-Jordan, H., Cheng, M. & Anderson, M. T cells in the control of organ-specific autoimmunity. *J. Clin. Invest.* **125**, 2250–2260 (2015).
6. Zou, W. Regulatory T cells, tumour immunity and immunotherapy. *Nature Rev. Immunol.* **6**, 295–307 (2006).
7. Nishikawa, H. & Sakaguchi, S. Regulatory T cells in tumor immunity. *Int. J. Cancer* **127**, 759–767 (2010).
8. Quezada, S. A., Peggs, K. S., Simpson, T. R. & Allison, J. P. Shifting the equilibrium in cancer immunoeediting: from tumor tolerance to eradication. *Immunol. Rev.* **241**, 104–118 (2011).

9. Huehn, J. *et al.* Developmental stage, phenotype, and migration distinguish naive- and effector/memory-like CD4⁺ regulatory T cells. *J. Exp. Med.* **199**, 303–313 (2004).
10. Miyara, M. *et al.* Functional delineation and differentiation dynamics of human CD4⁺ T cells expressing the FoxP3 transcription factor. *Immunity* **30**, 899–911 (2009).
11. Sugiyama, D. *et al.* Anti-CCR4 mAb selectively depletes effector-type FoxP3⁺CD4⁺ regulatory T cells, evoking antitumor immune responses in humans. *Proc. Natl Acad. Sci. USA* **110**, 17945–17950 (2013).
12. Ouyang, W. *et al.* Novel Foxo1-dependent transcriptional programs control T_{reg} cell function. *Nature* **491**, 554–559 (2012).
13. Kerdiles, Y. M. *et al.* Foxo transcription factors control regulatory T cell development and function. *Immunity* **33**, 890–904 (2010).
14. Smigiel, K. S. *et al.* CCR7 provides localized access to IL-2 and defines homeostatically distinct regulatory T cell subsets. *J. Exp. Med.* **211**, 121–136 (2014).
15. Mueller, S. N., Gebhardt, T., Carbone, F. R. & Heath, W. R. Memory T cell subsets, migration patterns, and tissue residence. *Annu. Rev. Immunol.* **31**, 137–161 (2013).
16. Hedrick, S. M., Hess Michelini, R., Doedens, A. L., Goldrath, A. W. & Stone, E. L. FOXO transcription factors throughout T cell biology. *Nature Rev. Immunol.* **12**, 649–661 (2012).
17. Ouyang, W. & Li, M. O. Foxo: in command of T lymphocyte homeostasis and tolerance. *Trends Immunol.* **32**, 26–33 (2011).
18. Luo, C. T. & Li, M. O. Transcriptional control of regulatory T cell development and function. *Trends Immunol.* **34**, 531–539 (2013).
19. Fabre, S. *et al.* FOXO1 regulates L-Selectin and a network of human T cell homing molecules downstream of phosphatidylinositol 3-kinase. *J. Immunol.* **181**, 2980–2989 (2008).
20. Kerdiles, Y. M. *et al.* Foxo1 links homing and survival of naive T cells by regulating L-selectin, CCR7 and interleukin 7 receptor. *Nature Immunol.* **10**, 176–184 (2009).
21. Ouyang, W., Beckett, O., Flavell, R. A. & Li, M. O. An essential role of the Forkhead-box transcription factor Foxo1 in control of T cell homeostasis and tolerance. *Immunity* **30**, 358–371 (2009).
22. Curiel, T. J. *et al.* Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nature Med.* **10**, 942–949 (2004).
23. Shimizu, J., Yamazaki, S. & Sakaguchi, S. Induction of tumor immunity by removing CD25⁺CD4⁺ T cells: a common basis between tumor immunity and autoimmunity. *J. Immunol.* **163**, 5211–5218 (1999).
24. Bos, P. D., Plitas, G., Rudra, D., Lee, S. Y. & Rudensky, A. Y. Transient regulatory T cell ablation deters oncogene-driven breast cancer and enhances radiotherapy. *J. Exp. Med.* **210**, 2435–2466 (2013).
25. Franklin, R. A. *et al.* The cellular and molecular origin of tumor-associated macrophages. *Science* **344**, 921–925 (2014).
26. Stringer, J. R. *et al.* Modeling variation in tumors in vivo. *Proc. Natl Acad. Sci. USA* **102**, 2408–2413 (2005).
27. Bauer, C. A. *et al.* Dynamic Treg interactions with intratumoral APCs promote local CTL dysfunction. *J. Clin. Invest.* **124**, 2425–2440 (2014).
28. Levine, A. G., Arvey, A., Jin, W. & Rudensky, A. Y. Continuous requirement for the TCR in regulatory T cell function. *Nature Immunol.* **15**, 1070–1078 (2014).
29. Vahl, J. C. *et al.* Continuous T cell receptor signals maintain a functional regulatory T cell pool. *Immunity* **41**, 722–736 (2014).
30. Ali, K. *et al.* Inactivation of PI(3)K p110 δ breaks regulatory T-cell-mediated immune tolerance to cancer. *Nature* **510**, 407–411 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Rudensky for the Foxp3^{Cre} mouse strain and S. Abrams for the AT-3 cell line. We also thank the Li laboratory for discussions. This work was supported by the National Institutes of Health (RO1 AI102888-01A1 to M.O.L.), Ludwig Center for Cancer Immunology (M.O.L.), and the Memorial Sloan Kettering Cancer Center Support Grant/Core Grant (P30 CA008748).

Author Contributions C.T.L. and M.O.L. were involved in all aspects of this study including planning and performing experiments, analysing and interpreting data, and writing the manuscript. W.L. processed and analysed next-generation sequencing data, and wrote the manuscript. S.D. and A.T. participated in mouse colony management and experiments including parabiosis. M.O.L. oversaw all the work performed.

Author Information RNA-sequencing data are available at the Gene Expression Omnibus (GEO) repository under accession number GSE74957. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.O.L. (lim@mskcc.org).

METHODS

Mice. The Foxo1CA (*Foxo1^{AAA}* knock-in) and Foxo1tag-GFP (*Foxo1^{tag}*) mouse models, as well as MMTV-PyMT on C57BL/6 background were previously described^{12,25}. C57BL/6, CD45.1⁺ and CD8^{-/-} mice were purchased from Jackson Laboratory. Mice with T_{reg}-cell-specific expression of Foxo1CA were generated by crossing *Foxo1^{AAA}* with *Foxp3^{Cre}* mice³¹. *Foxp3^{Cre}* mice express yellow fluorescent protein (YFP) from the *Foxp3* locus, and were used as a reporter for sorting experiments. To mark T_{reg} cells with red fluorescent protein (RFP) in the Foxo1tag-GFP experiment, *Foxo1^{tag}* mice were bred with *Foxp3-IRES-RFP* mice³². In all experiments, littermate controls were used when possible. Unless mentioned otherwise, both male and female mice were included. All mice were maintained under specific pathogen-free conditions, and animal experimentation was conducted in accordance with procedures approved by the Institutional Animal Care and Use Committee of Memorial Sloan Kettering Cancer Center. Investigators were not blinded to group allocation and outcome assessment. No statistical methods were used to predetermine sample size and the experiments were not randomized.

Parabiosis. Parabiosis and separation were done as reported³³ with 6–8-week-old congenically marked female C57BL/6 mice that were matched for body weight. In brief, matching skin incisions were made from the elbow to the knee of each mouse. Forelimb and hindlimb connections were made with sutures and skin incisions were closed using woundclips. For separation, connections between parabionts were disrupted and skin incisions were closed using woundclips. Parabionts were maintained for 4 weeks before surgical separation. Separated parabiont mice were analysed up to 18 weeks after surgery.

Tumour models. MMTV-PyMT spontaneous tumour model was previously described³⁴. AT-3 model: 8–10-week-old wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+), *CD8^{-/-}*, or *Foxp3^{Cre}Foxo1CA/+CD8^{-/-}* (CA/+CD8^{-/-}) mice were injected with AT-3 mammary tumour³⁵ (2×10^5 cells into mammary fat pad). B16 model: 8–10-week-old wild-type or CA/+ mice were injected with B16.F10 melanoma (1.25×10^5 cells subcutaneously). For all tumour models, tumours were measured regularly with a caliper. Tumour volume was calculated using the equation ($L \times W^2$) $\times 0.52$, in which L denotes length, and W denotes width. The maximal tumour burden was 3,000 mm³, and in none of the experiments was the limit exceeded. For PyMT, individual tumour volumes were added together to calculate total tumour burden. Tumour bearing mice were euthanized at 22–24 weeks old for the PyMT model, 30–35 days after inoculation for the AT-3 model and 20–24 days after inoculation for the B16 model. In all flow cytometry experiments, tumours of similar sizes were used for comparison.

Cell isolation. After whole-body perfusion with 50 ml of heparinized PBS, lymphocytes were isolated as follows. Single-cell suspensions were prepared from spleens and peripheral (axillary, brachial and inguinal) lymph nodes by tissue disruption with glass slides. To isolate cells from the liver, tissues were finely minced and digested with 1 mg ml⁻¹ collagenase D (Worthington) for 30 min at 37°C. For lamina propria lymphocytes isolation, colon was dissected and washed in HBSS. Intestinal pieces were stirred in 1 mM dithiothreitol (DTT) in HBSS to release intraepithelial lymphocytes. The remaining intestinal tissues were finely minced and digested with RPMI plus 5% FBS and 1 mg ml⁻¹ collagenase D for 30 min at 37°C. For tumour-infiltrating immune cell isolation, tumour tissues were prepared by mechanical disruption followed by 1 h treatment with 280 U ml⁻¹ collagenase type 3 (Worthington) and 4 µg ml⁻¹ DNase I (Sigma) at 37°C. After the digestion steps, cells isolated from the liver, lamina propria and tumour were filtered through a 70-µm cell strainer, layered in a 44% and 66% Percoll gradient (Sigma), and centrifuged at 1,600g for 30 min without brake. Cells at the interface were collected and analysed by flow cytometry.

Flow cytometry. Fluorochrome-conjugated, biotinylated antibodies against CD45.1 (clone 104), CD45.2 (A20), TCR-β (H57-595), CD4 (RM4-5), CD8 (17A2), CD44 (IM7), CD62L (MEL-14), CD69 (H1.2F3), CCR7 (4B12), Foxp3 (FJK-16 s), ICOS (C398.4A), IFN-γ (XMG1.2), IL-4 (11B11) and NK1.1 (PK136) were purchased from eBioscience. Antibodies against CD49a (Ha31/8) and IL-17a (TC11-18H10.1) were purchased from BD Biosciences. Anti-GzmB (GB11) was purchased from Invitrogen. Purified antibodies against Bim (C34C5), p-Akt(S473) (736E11), p-Akt(T308) (C31E5E), p-Foxo1(T24) and p-S6(S240) were purchased from Cell Signaling. All antibodies were tested with their respective isotype controls. Cell surface staining was performed by incubating cells with specific antibodies for 30 min on ice in the presence of 2.4G2 monoclonal antibody to block FcγR binding. CCR7 staining was incubated in 37°C for 30 min before cell surface staining. Foxp3, Bim, GzmB, IFN-γ, IL-4 and IL-17a staining was carried out using the intracellular transcription factor or cytokine staining kits from Tonbo or BD Biosciences. Phosphorylation staining was performed using BD phospho-protein kit. Secondary antibodies with fluorochrome-conjugation were used for the staining of purified antibodies. To determine cytokine expression, isolated cells were stimulated with 50 ng ml⁻¹ phorbol 12-myristate 13-acetate (Sigma), 1 mM ionomycin (Sigma) and GolgiStop

(BD Biosciences) for 4 h before staining. Incorporation of EdU was measured using the Click-iT EdU flow cytometry assay kit according to the manufacturer's instructions (Invitrogen). Mice were injected intraperitoneally with 50 µg g⁻¹ body weight of EdU and euthanized 18 h later. For all stains, dead cells were excluded from analysis by means of Live/Dead Fixable Dye (Invitrogen), DAPI or propidium iodide stain. All samples were acquired and analysed with LSRII flow cytometer (Becton Dickinson) and FlowJo software (TreeStar).

In vivo T-cell labelling. Anti-CD4-biotin (RM4-4, BioLegend) (2 µg) was injected intravenously, and mice were euthanized 5 min after injection. Splenocytes were prepared for flow cytometric analysis as described above. Streptavidin conjugated with fluorophore was used as the secondary labelling for anti-CD4-biotin, and a non-competing clone of CD4 antibody (RM4-5, eBioscience) was used to stain total CD4⁺ cells.

In vitro T-cell culture. Akt inhibitor experiment: CD4⁺Foxp3⁺CD62L^{hi}CD44^{lo} rT_{reg} cells were purified from spleen and lymph nodes of *Foxp3^{YFP/Cre}* mice by flow cytometry sorting (BD FACS Aria) at Flow Cytometry Core Facility of Memorial Sloan Kettering Cancer Center (MSKCC). rT_{reg} cells were cultured with plate-bound anti-CD3 (coated overnight, 5 µg ml⁻¹), soluble anti-CD28 (2 µg ml⁻¹) and IL-2 (200 U ml⁻¹) for 3 days. Akt inhibitor (MK-2206 2HCl, Selleckchem, 2 µM final concentration) or DMSO solvent control was added to the culture. rT_{reg} cell activation experiments: CD4⁺Foxp3⁺CD62L^{hi}CD44^{lo}CD69⁻ICOS⁻ rT_{reg} cells were sorted out from *Foxp3^{Cre}*, *Foxp3^{Cre}Foxo1CA/+* or *Foxp3^{Cre}Foxo1CA/Foxo1CA* mice, and labelled with CellTrace Violet (Invitrogen). Cells were cultured with plate-bound anti-CD3 (coated overnight, 5 µg ml⁻¹), soluble anti-CD28 (2 µg ml⁻¹) and IL-2 (200 U ml⁻¹) for 3 days.

In vitro T_{reg} cell suppression. CD4⁺CD25⁻CD62L^{hi}CD44^{lo} conventional T cells purified by flow cytometry sorting were labelled with CellTrace Violet and used as responder cells. Responder T cells (5×10^4) were cultured for 72 h with irradiated splenocytes (1×10^5) and anti-CD3 (2 µg ml⁻¹) in the presence or absence of various numbers of T_{reg} cells. T_{reg} cells (CD4⁺Foxp3⁺YFP⁺) were isolated by FACS sorting from *Foxp3^{Cre}*, *Foxp3^{Cre}Foxo1CA/+* or *Foxp3^{Cre}Foxo1CA/Foxo1CA* mice.

Gene-expression profiling. Splenic rT_{reg} (CD4⁺Foxp3⁺YFP⁺CD62L^{hi}CD44^{lo}) and aT_{reg} (CD4⁺Foxp3⁺YFP⁺CD62L^{lo}CD44^{hi}) cells were isolated from *Foxp3^{YFP/Cre}* mice by FACS sorting. RNA was prepared with the miRNeasy Mini Kit according to the manufacturer's instructions (Qiagen). Complementary DNA (cDNA) libraries were amplified using the SMARTer RACE Amplification Kit (Clontech), and were sequenced in replicate using 50-base-pair paired-end reads at Genomics Core Laboratory of MSKCC. Ribosomal RNA reads were quantified and filtered using the short-read aligner, Bowtie v2.1.0 (ref. 34). The remaining reads were aligned to the mouse genome (mm10) using the STAR v2.3.1 short-read aligner³⁶. Additional quality control was performed using RSeQC v2.3.7 (ref. 37). Gene abundance was quantified by featureCounts³⁸ via the Subread analysis suite v1.4.3 (ref. 39). Differential gene expression was estimated using the DESeq2 R package with gene annotations curated in GENCODE version 2 for mouse reference genome GRCm38 (Ensembl 74)⁴⁰.

Quantitative PCR. RNA extraction was performed using RNeasy columns (Qiagen) and cDNA was generated using QuantiTect Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions. RT² SYBR Green kit (Qiagen) and Stratagene Mx3500 (Agilent) were used in the qPCR experiments. mRNA levels of *Lamc1*, *Nid2*, *Mmp9*, *Cd69*, *Egr2*, *Il1r2* and *Actb* were determined by the primers listed below. The mRNA amounts were normalized to those of *Actb*.

Lamc1: 5'-GGTGGTCTGTTTCAGCCATT-3' and 5'-TGCCACAAAATCTCAGCTTG-3'; *Nid2*: 5'-TGGATATGGCCAAGGAGAA-3' and 5'-CACCAGGACAGTTTCCATT-3'; *Mmp9*: 5'-ACCTTCCAGTAGGGGCAACT-3' and 5'-TGAATCAGCTGGCTTTTGTG-3'; *Cd69*: 5'-TCCGTGGACCACTTGAGAGT-3' and 5'-ATACCTGGTCCATGGTCCTT-3'; *Egr2*: 5'-AGGCCGTAGACAAAATCCCA-3' and 5'-TGATCATGCCATCTCCCGC-3'; *Il1r2*: 5'-TGGTGACACAGGAAAGGTT-3' and 5'-TGGATGTCGAGTGCAGTAGAGGT-3'; *Actb*: 5'-TTGCTGACAGGATGCAGAAG-3' and 5'-ACATCTGCTGGAAGGTGGAC-3'.

Immunofluorescence microscopy. Foxo1 localization experiment: spleens from wild-type mice were immediately disrupted using glass slides into Cytofix/Cytoperm buffer (BD). After incubation for 30 min at room temperature, the cells were washed, resuspended in 90% methanol and incubated on ice for 30 min. After an additional wash, cells were stained for surface and intracellular antigens, including CD4, CD44 and Foxp3, for 45 min at room temperature in the dark. rT_{reg} (CD4⁺Foxp3⁺CD44^{lo}) and aT_{reg} (CD4⁺Foxp3⁺CD44^{hi}) cells were then purified by FACS sorting, and spun to glass slides using Cytospin centrifuge (Thermo Scientific) at 1,200 r.p.m. for 5 min. Cells on the glass slides were stained with 1:150 diluted anti-Foxo1 (C29H4, Cell Signaling), followed by fluorophore-conjugated secondary antibody staining. Slides were mounted with gold anti-fading mounting buffer (Invitrogen). Images were acquired with a Leica TCS SP5-II confocal microscope at Molecular Cytology Core of MSKCC. For quantitative analysis, five

fields were selected randomly and total cells in the field were manually counted and grouped with Volocity software (PerkinElmer Inc.), on the basis of their Foxo1 nuclear or cytosolic localization.

For HA–Foxo1 staining, CD4⁺Foxp3–YFP⁺ cells were purified from *Foxp3^{Cre}*, *Foxp3^{Cre}Foxo1CA/+* or *Foxp3^{Cre}Foxo1CA/Foxo1CA* mice by FACS sorting, and were spun to glass slides using Cytospin centrifuge (Thermo Scientific) at 500 r.p.m. for 1 min. After fixation with 4% paraformaldehyde, cells were permeabilized with Foxp3 fixation/permeabilization buffer (eBioscience) according to the manufacturer's instructions. Cells were then blocked with permeabilization buffer and 3% BSA, and incubated with 1:500 diluted anti-HA (C29F4, Cell Signaling) and 1:150 diluted anti-Foxp3 (FJK-16s, eBioscience), followed by fluorophore-conjugated secondary antibodies in permeabilization buffer and 1% BSA.

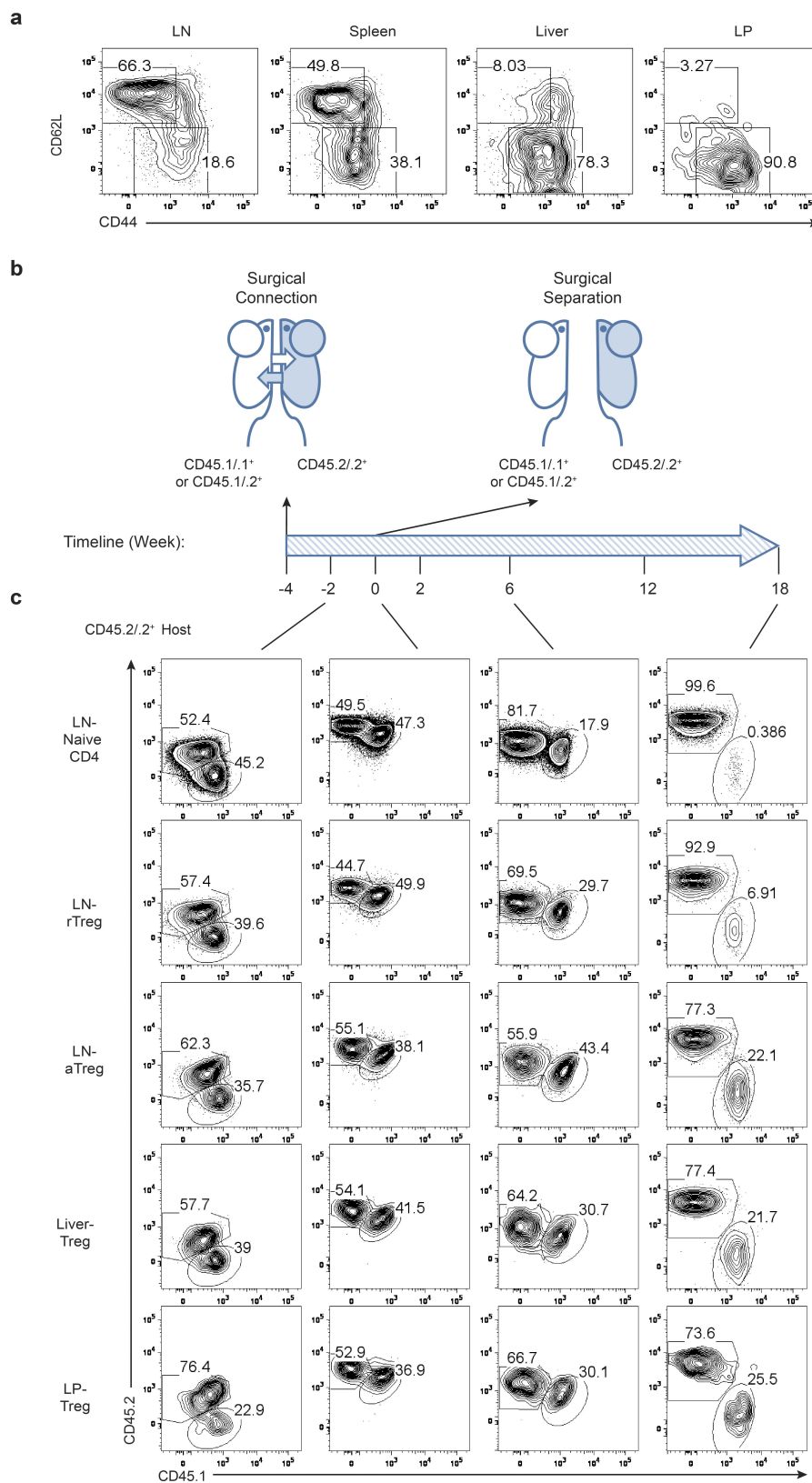
Immunoblotting. T_{reg} cells (CD4⁺Foxp3–YFP⁺) were isolated from *Foxp3^{Cre}*, *Foxp3^{Cre}Foxo1CA/+* or *Foxp3^{Cre}Foxo1CA/Foxo1CA* mice by FACS sorting. Total protein extracts were dissolved in SDS sample buffer, separated on 12% SDS–PAGE gels and transferred to polyvinylidene difluoride membrane (Millipore). The membranes were probed with anti-HA (6E2, Cell Signaling) and anti-β-actin (AC-15, Sigma), and visualized with the Immobilon Western Chemiluminescent HRP Substrate (Millipore).

Histopathology. Liver and colon tissues were fixed in Safefix II (Protocol) and embedded in paraffin. Five-millimetre sections were stained with haematoxylin and eosin. The following grades were used to evaluate the colon pathology: 0, normal colonic crypt architecture; 1, mild inflammation: slight epithelial cell hyperplasia and increased numbers of leukocytes in the mucosa; 2, moderate colitis: pronounced epithelial hyperplasia, marked leukocyte infiltration, and decreased numbers of goblet cells; 3, severe colitis: marked epithelial hyperplasia with extensive leukocyte infiltration, substantial depletion of goblet cells, occasional ulceration, or cryptic abscesses; 4, very severe colitis: marked epithelial hyperplasia with extensive transmural leukocyte infiltration, severe depletion of goblet cells, many crypt abscesses and severe ulceration.

Serum ALT. Blood was collected immediately after mice were euthanized and was stored at room temperature for 1 h. The samples were then centrifuged for 15 min at 1,800g, and the supernatant was obtained as serum. ALT activity was determined according to manufacturer's instructions (Sigma–Aldrich), using SpectraMax M5 plate reader (Molecular Devices).

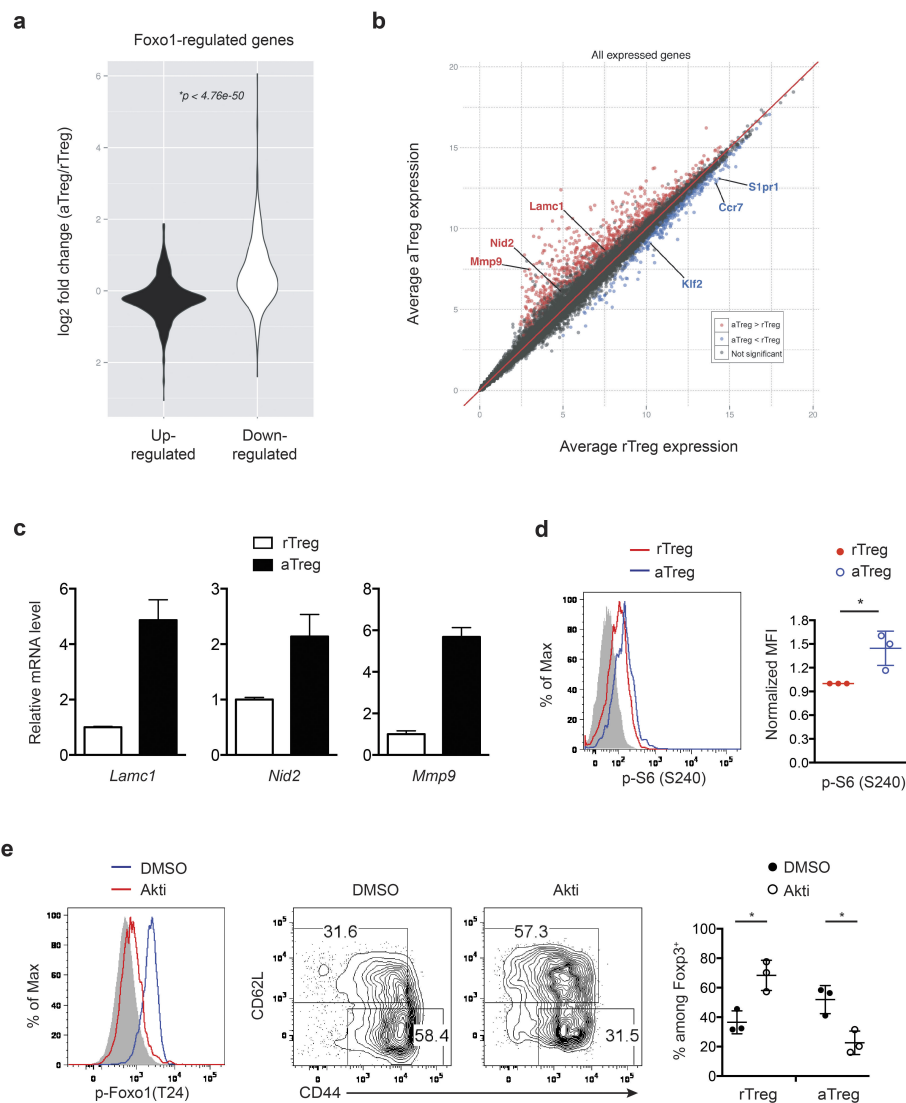
Statistical analysis. All data are presented as the mean ± s.e.m. Comparisons between groups were analysed using unpaired or paired Student's *t*-tests or ANOVA, as appropriate. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

31. Rubtsov, Y. P. *et al.* Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28**, 546–558 (2008).
32. Wan, Y. Y., Chi, H., Xie, M., Schneider, M. D. & Flavell, R. A. The kinase TAK1 integrates antigen and cytokine receptor signaling for T cell development, survival and function. *Nature Immunol.* **7**, 851–858 (2006).
33. Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L. & Weissman, I. L. Physiological migration of hematopoietic stem and progenitor cells. *Science* **294**, 1933–1936 (2001).
34. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
35. Stewart, T. J. & Abrams, S. I. Altered immune function during long-term host-tumor interactions can be modulated to retard autochthonous neoplastic growth. *J. Immunol.* **179**, 2851–2859 (2007).
36. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
37. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
38. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
39. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
40. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** Suppl 1, 1–9 (2006).



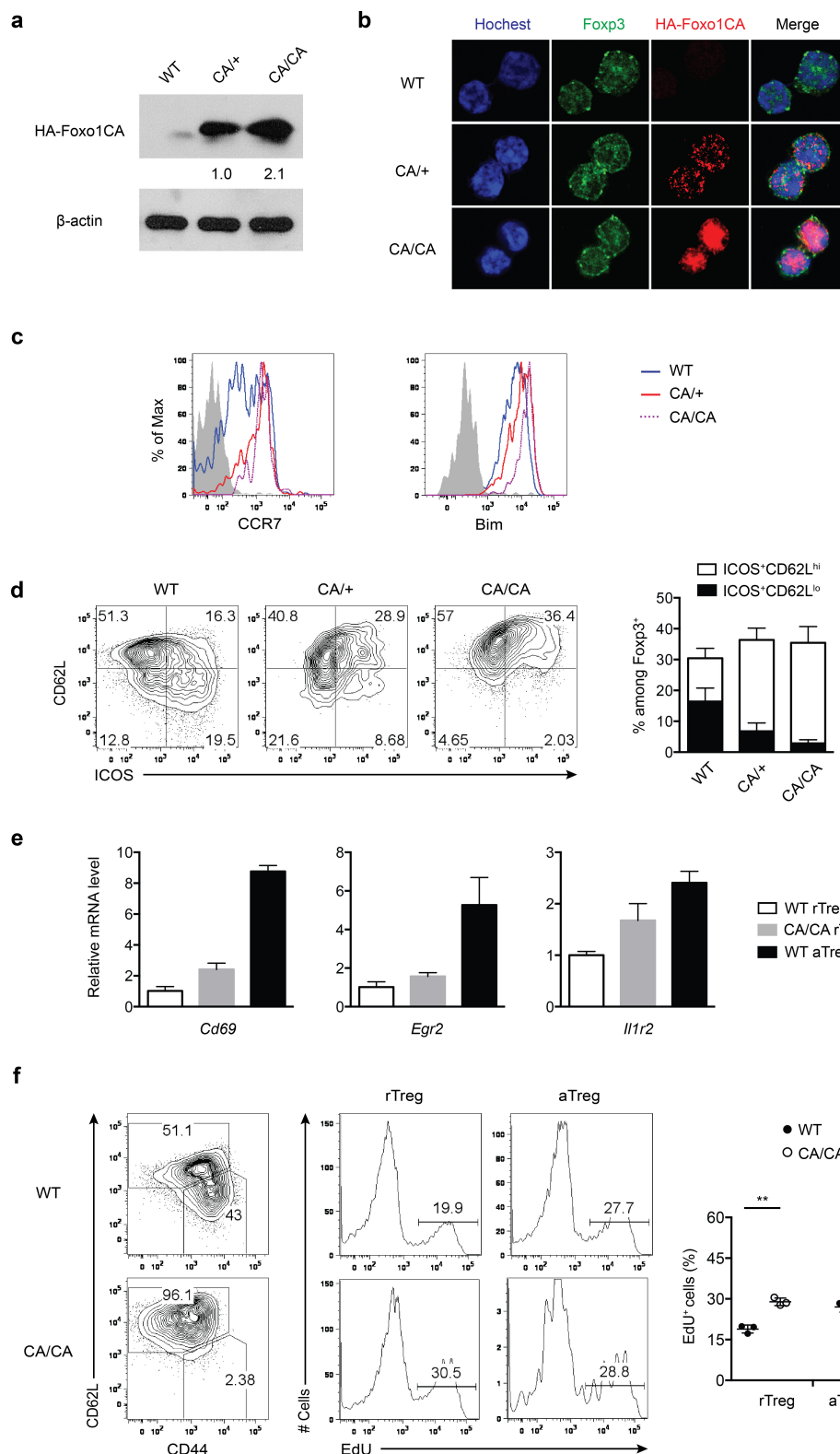
Extended Data Figure 1 | Parabiotic analysis of T_{reg} cells from different organs. **a**, Flow cytometric analysis of CD44 and CD62L in T_{reg} cells (CD4⁺Foxp3⁺) from lymph node, spleen, liver and colon lamina propria of C57BL/6 mice. Percentages of rT_{reg} and aT_{reg} cells are shown. A sizable population of intermediate phenotype (CD62L^{hi}CD44^{hi}) T_{reg} cells was also present in the liver. **b**, Graphical representation of the parabiosis experiments. Congenically mismatched C57BL/6 mice were surgically

connected (time point -4). Parabionts were analysed 2 or 4 weeks after surgery (time point -2 or 0). In separation experiments, parabiotic mice that had been connected for 4 weeks were surgically disconnected from each other (time point 0). Separated mice were analysed 2, 6, 12 or 18 weeks after surgery (time point 2, 6, 12 or 18). **c**, Representative flow cytometric plots showing chimerism of naive CD4⁺ T cells or different T_{reg} cell subsets in a CD45.2/2⁺ parabiont at various time points.



Extended Data Figure 2 | aTreg cell differentiation is associated with Akt-triggered suppression of Foxo1. **a**, Gene expression comparison of Foxo1-regulated genes in splenic aTreg versus rTreg cells. Foxo1-regulated genes were defined by the following criteria: (1) differentially expressed between wild-type and Foxo1-knockout Treg cells; (2) the expression was corrected by expression of a constitutively active mutant of Foxo1 (Foxo1CA). **b**, Normalized expression of all transcripts isolated from splenic aTreg cells were plotted against transcripts from splenic rTreg cells. Some of the Foxo1 direct target genes were highlighted. **c**, Comparison of *Lamc1*, *Nid2*, *Mmp9* mRNA levels in rTreg versus aTreg subsets. **d**, Flow cytometric analysis of phosphorylated S6 ribosomal

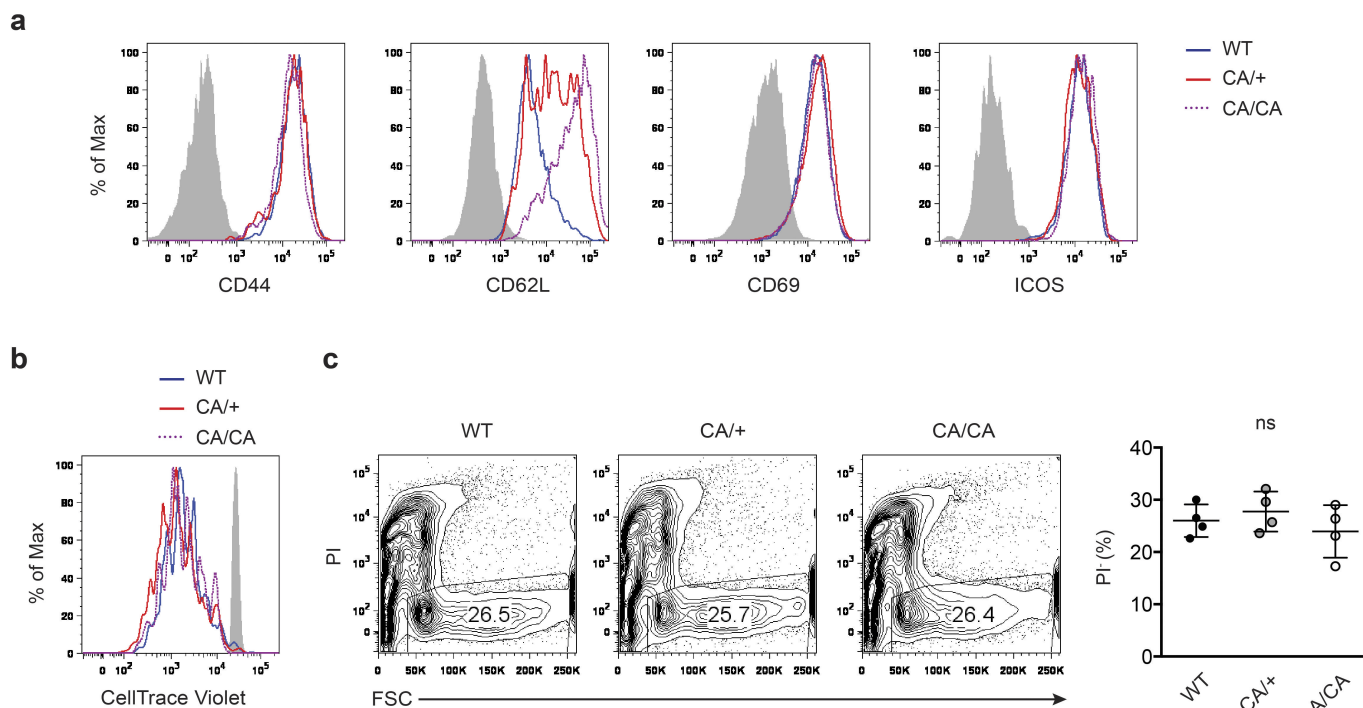
protein in splenic aTreg and rTreg cells. Quantification shows mean fluorescence intensity normalized to rTreg cells. $n = 3$. $*P < 0.05$ (paired t -test). **e**, aTreg cell differentiation experiments *in vitro*. rTreg cells ($CD4^+Foxp3^+CD62L^{hi}CD44^{lo}$) from spleen and lymph nodes of Foxp3-YFP reporter mice were purified by flow cytometric sorting, and were stimulated with anti-CD3/CD28 and IL-2 for 3 days. Akt inhibitor (Akti, MK-2206) or DMSO solvent control were added to the culture. Phosphorylated Foxo1, and CD44 and CD62L levels were determined by flow cytometry. Quantification shows percentages of rTreg and aTreg cells in DMSO control or Akt inhibitor group. $n = 3$. $*P < 0.05$ (unpaired t -test). Data are mean \pm s.e.m.



Extended Data Figure 3 | Foxo1 hyperactivation in T_{reg} cells does not affect expression of activation markers or proliferation.

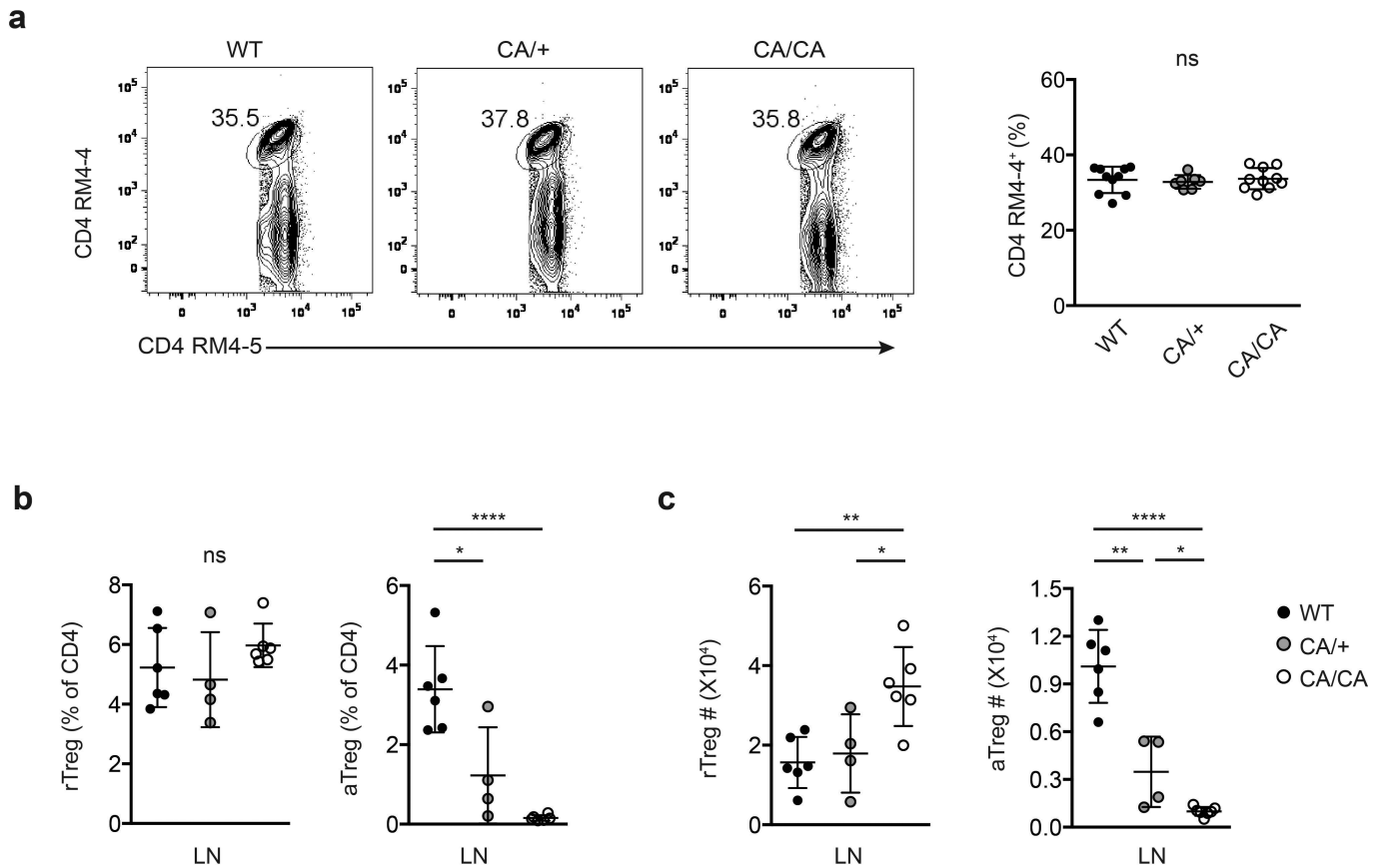
a, b, T_{reg} cells (CD4⁺Foxp3⁺) from spleens and lymph nodes of wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice were purified by flow cytometric sorting. **a**, Haemagglutinin (HA)-tagged Foxo1CA protein in the whole-cell lysate was measured by immunoblotting with anti-HA. **b**, Subcellular localization of HA-Foxo1CA was determined by immunofluorescence staining with anti-HA. **c**, Flow cytometric analysis of Foxo1-activated target genes CCR7 and Bim in lymph node T_{reg} cells from 9–12-day-old wild-type, CA/+ and CA/CA mice. Grey shaded lines represent isotype controls. **d**, Flow cytometric

analysis of ICOS and CD62L expression in lymph node T_{reg} cells from 9–12-day-old wild-type, CA/+ and CA/CA mice. Bar graph shows fractions of ICOS⁺CD62L^{hi}, ICOS⁺CD62L^{lo} subsets among T_{reg} cells. *n* = 4. **e**, CD62L^{hi}CD44^{hi} rT_{reg} cells from wild-type or CA/CA mice, or CD62L^{lo}CD44^{hi} aT_{reg} cells from wild-type mice were isolated by flow cytometric sorting, and the expression of *Cd69*, *Egr2* and *Il1r2* mRNA was determined by reverse transcriptase quantitative PCR (RT-qPCR). **f**, EdU was administrated intraperitoneally into 9–10-day-old wild-type or CA/CA mice. EdU incorporation in lymph node rT_{reg} and aT_{reg} cells was analysed 18 h after injection. *n* = 3. ***P* < 0.01 (unpaired *t*-test). Data are mean ± s.e.m.



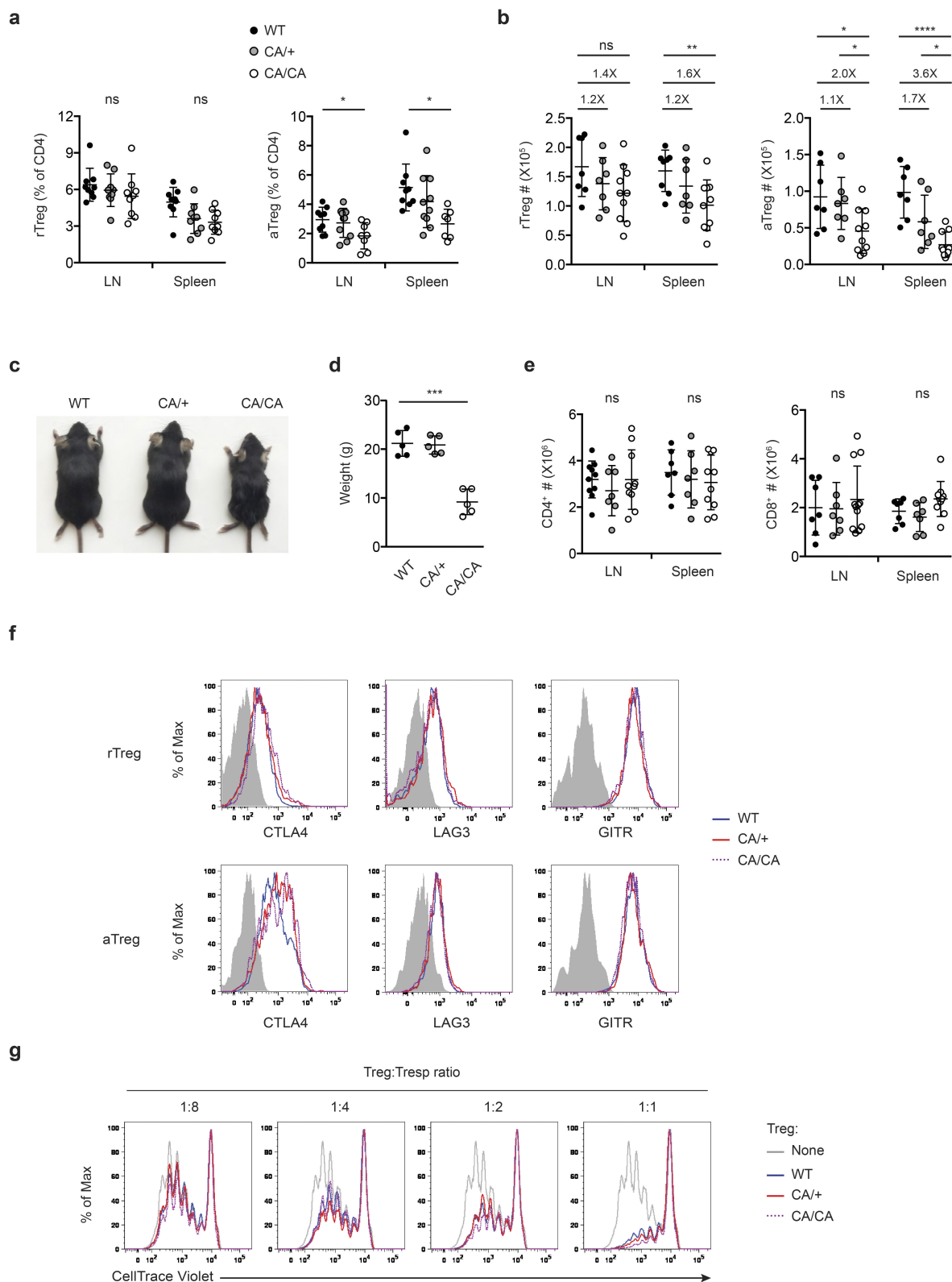
Extended Data Figure 4 | CA-expressing T_{reg} cells show intact activation, proliferation and survival *in vitro*. CD4⁺Foxp3⁺CD62L^{hi}CD44^{lo}CD69⁻ICOS⁻ rT_{reg} cells were isolated from wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice, and were stimulated with anti-CD3, anti-CD28 and IL-2 for 3 days. **a**, CD44, CD62L, CD69 and ICOS expression was determined by flow cytometry.

Grey shaded lines represent isotype controls. **b**, Cells were stained with CellTrace Violet dye before culture, and cell division was tracked by dilution of the dye. Grey shaded line represents undivided cells. **c**, Cell death was measured by propidium iodide (PI) incorporation. Percentages of live cell fraction (PI⁻) were compared. $n = 4$. NS, not significant (unpaired *t*-test). Data are mean \pm s.e.m.



Extended Data Figure 5 | Foxo1 hyperactivation depletes aT_{reg} cells in mice. **a**, *In vivo* labelling experiment: anti-CD4-biotin (RM4-4 clone) was injected intravenously into wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice, and spleens were collected 5 min after injection. A non-competing clone of anti-CD4 (RM4-5) was used in subsequent flow cytometric analysis to label total CD4⁺ T cells. Representative flow cytometric plots from gated conventional CD4⁺

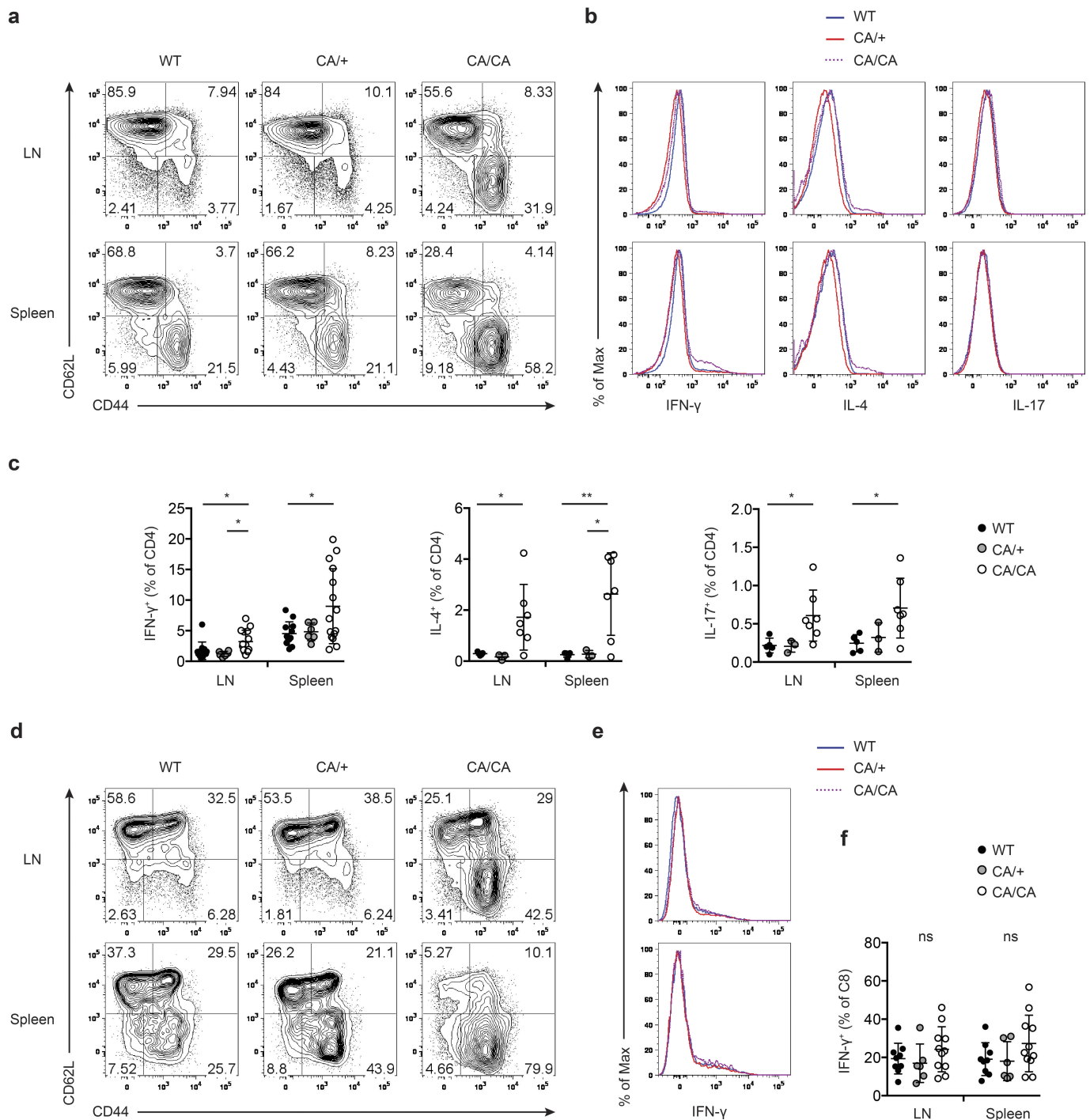
T cells (CD4 RM4-5⁺Foxp3⁻) are shown. Quantification shows percentage of RM4-4⁺ among conventional CD4⁺ T cells. *n* = 10 (WT, CA/CA); *n* = 7 (CA/+). **b**, **c**, The frequencies of lymph node rT_{reg} or aT_{reg} cells among CD4⁺ T cells of 9–12-day-old wild-type, CA/+ or CA/CA mice. *n* = 4–6. **c**, The numbers of lymph node rT_{reg} or aT_{reg} cells of 9–12-day-old wild-type, CA/+ or CA/CA mice. *n* = 6 (WT, CA/CA); *n* = 4 (CA/+). Unpaired *t*-test. Data are mean ± s.e.m.



Extended Data Figure 6 | Foxo1 hyperactivation preferentially impairs aTreg cells, but does not affect Treg cell suppressive function *in vitro*.

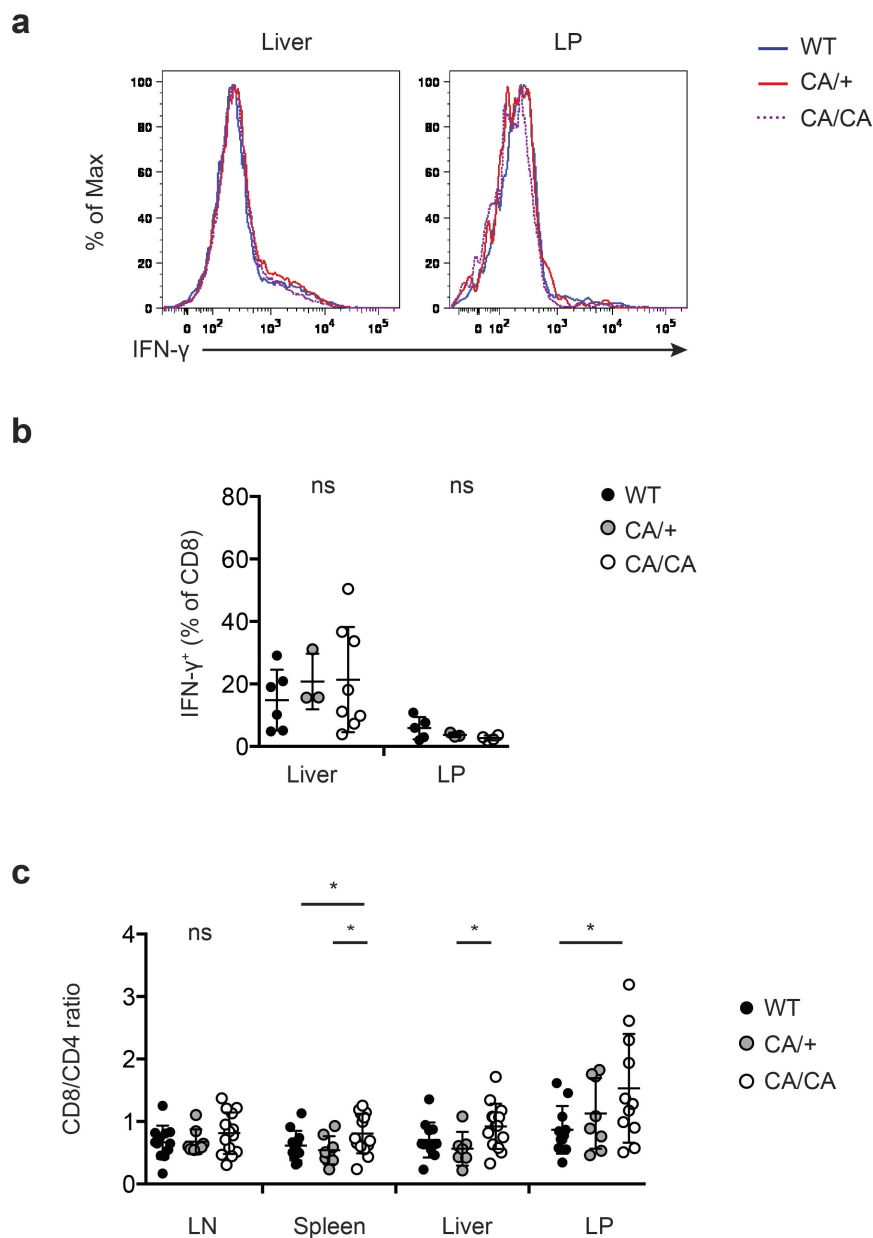
a, The frequencies of lymph node and splenic rTreg or aTreg cells among CD4⁺ T cells of 4–6-week-old wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice. *n* = 10. **b**, The numbers of lymph node rTreg or aTreg cells of 4–6-week-old wild-type, CA/+ or CA/CA mice. *n* = 10. **c**, A representative picture of 5-week-old wild-type, CA/+ or CA/CA mice. **d**, Body weight of 4–6-week-old wild-type, CA/+ and CA/CA mice. *n* = 5. **e**, The numbers of CD4⁺ and CD8⁺ T cells from lymph node and spleen of 4–6-week-old wild-type, CA/+ and CA/CA

mice. *n* = 7 (WT, CA/+); *n* = 14 (CA/CA). **f**, Flow cytometric analysis of CTLA4, LAG3 and GITR expression in lymph node rTreg and aTreg cells from 4–6-week-old wild-type, CA/+ and CA/CA mice. Grey shaded lines represent isotype controls. **g**, Suppression of wild-type naive CD4⁺ T cells, labelled with CellTrace Violet (T_{resp}, responding T cells), by wild-type, CA/+ or CA/CA Treg cells. T_{resp} cell division was assessed by CellTrace Violet dilution at the indicated ratios of cell numbers between Treg and T_{resp} cells. Grey line represents conditions without Treg cell in culture. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (unpaired *t*-test). Data are mean ± s.e.m.



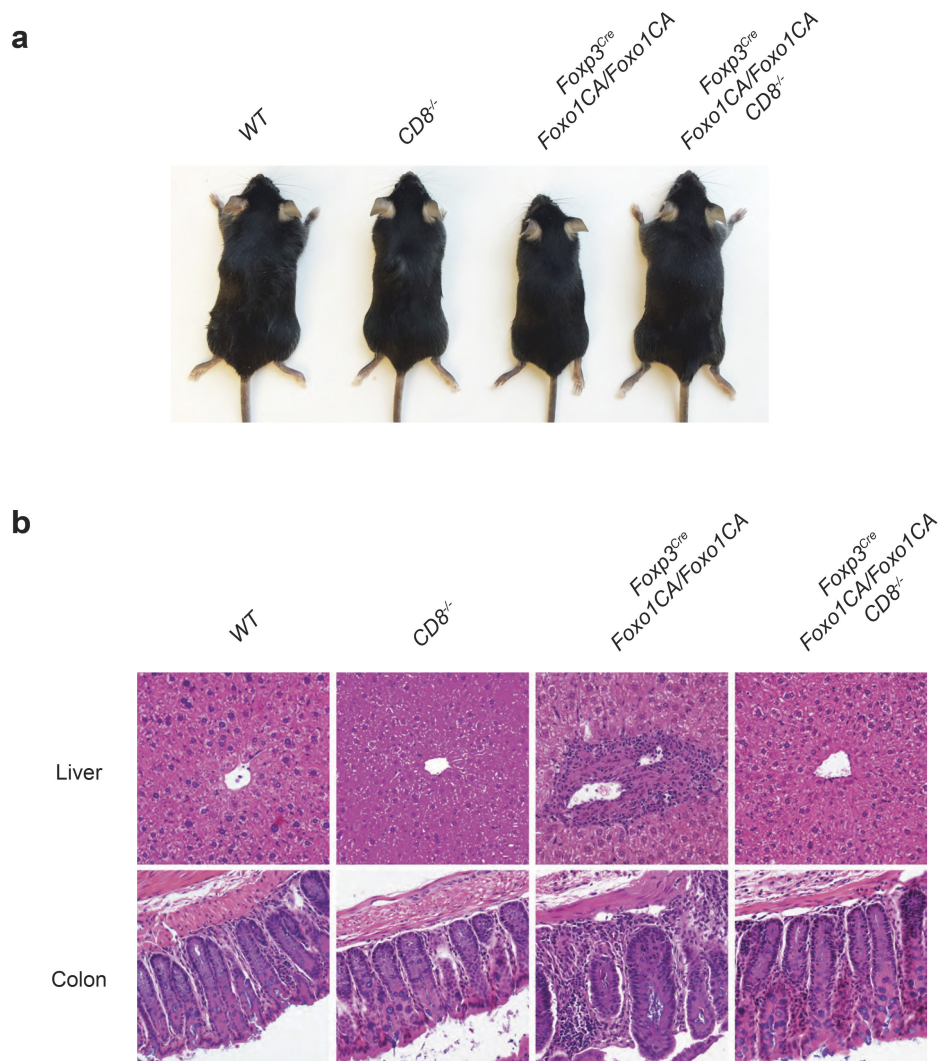
Extended Data Figure 7 | Modest increase of inflammatory cytokine production by CD4⁺ and CD8⁺ T cells in *Foxp3^{Cre}Foxo1CA/Foxo1CA* mice. **a, d, Flow cytometric analysis of CD44 and CD62L expression in CD4⁺Foxp3[−] conventional T cells (**a**) or CD8⁺ T cells (**d**) from lymph node and spleen of 4–6-week-old wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice. **b, c**, Representative histogram (**b**), and quantification (**c**) of cytokine (IFN- γ , IL-4, IL-17)**

expression by CD4⁺Foxp3[−] conventional T cells from lymph node and spleen of 4–6-week-old wild-type, CA/+ and CA/CA mice. $n = 14$ (WT, CA/CA for IFN- γ), $n = 7$ (WT, CA/CA for IL-4, IL-17); $n = 3$ (CA/+). **e, f**, Representative histogram (**e**), and quantification (**f**) of IFN- γ expression by CD8⁺ T cells from lymph node and spleen of 4–6-week-old wild-type, CA/+ and CA/CA mice. $n = 9$ (WT); $n = 6$ (CA/+); $n = 11$ (CA/CA). * $P < 0.05$; ** $P < 0.01$ (unpaired t -test). Data are mean \pm s.e.m.



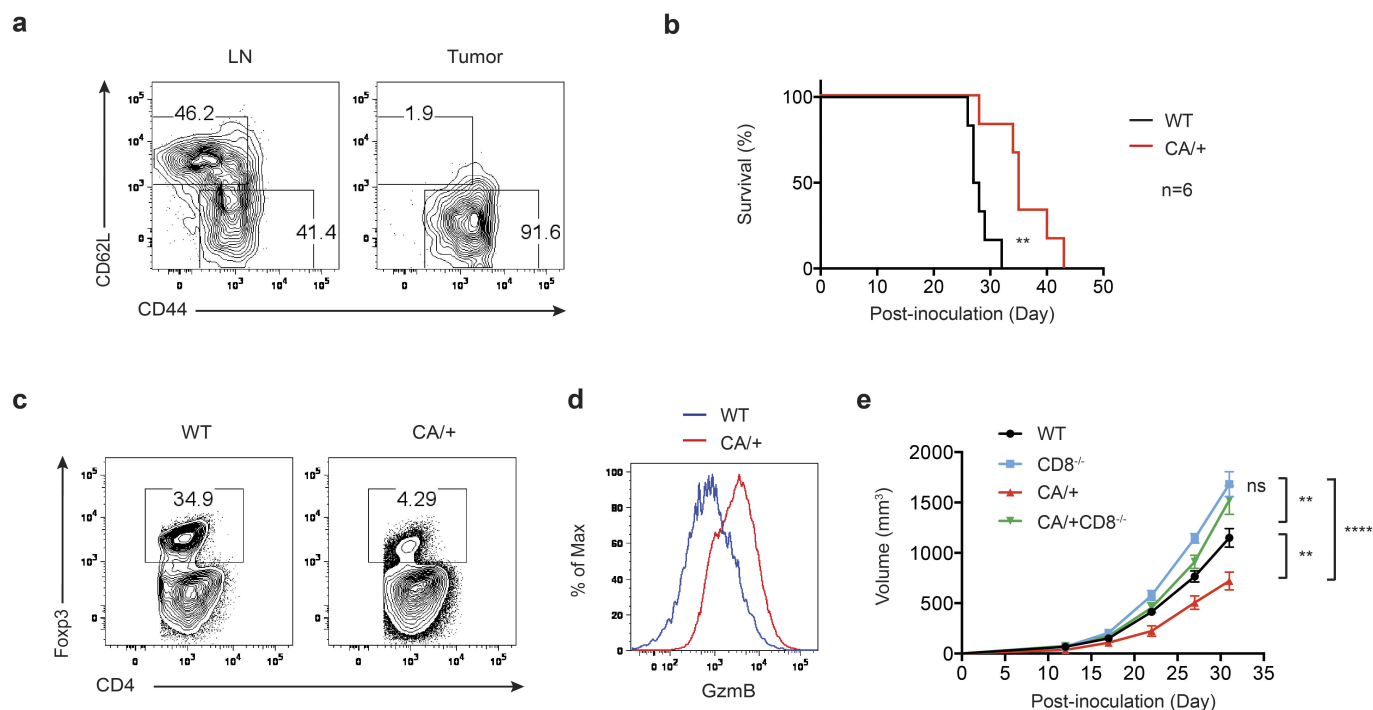
Extended Data Figure 8 | Characterization of CD8⁺ T cells from non-lymphoid tissues. **a, b,** Flow cytometric analysis (**a**), and quantification (**b**) of IFN- γ expression by CD8⁺ T cells from liver and colon lamina propria of 4–6-week-old wild-type, *Foxp3^{Cre}Foxo1CA/+* (CA/+) or *Foxp3^{Cre}Foxo1CA/Foxo1CA* (CA/CA) mice. $n = 6$ (WT); $n = 3$

(CA/+); $n = 8$ (CA/CA). **c,** CD8⁺ to CD4⁺ T-cell ratio among lymph node, spleen, liver and lamina propria T cells of 4–6-week-old wild-type, CA/+ and CA/CA mice. $n = 10$ (WT); $n = 8$ (CA/+); $n = 13$ (CA/CA). * $P < 0.05$ (unpaired *t*-test). Data are mean \pm s.e.m.



Extended Data Figure 9 | $CD8^{+}$ T-cell depletion rescues the lethal disease in $Foxp3^{Cre}Foxo1CA/Foxo1CA$ mice. a, A representative image of 5-week-old wild-type, $CD8^{-/-}$, $Foxp3^{Cre}Foxo1CA/Foxo1CA$ and

$Foxp3^{Cre}Foxo1CA/Foxo1CA CD8^{-/-}$ mice. **b,** Haematoxylin and eosin staining of liver and colon sections from 5-week-old wild-type, $CD8^{-/-}$, $Foxp3^{Cre}Foxo1CA/Foxo1CA$ and $Foxp3^{Cre}Foxo1CA/Foxo1CA CD8^{-/-}$ mice.



Extended Data Figure 10 | *Foxp3^{Cre}Foxo1CA/+* mice show enhanced anti-tumour immune responses. **a**, Flow cytometric analysis of CD44 and CD62L in T_{reg} cells ($CD4^{+}Foxp3^{+}$) from lymph node and tumour of PyMT tumour-bearing mice. **b–d**, Eight-to-ten-week-old wild-type or *Foxp3^{Cre}Foxo1CA/+* (CA/+) mice received orthotopic inoculation of PyMT-derived mammary tumour cells (AT-3). **b**, Survival of wild-type and CA/+ mice received tumour implantation. **c**, Foxp3 expression in

tumour-infiltrating $CD4^{+}$ T cells from day 30–35 AT-3 tumour. **d**, Flow cytometric analysis of GzmB expression in $CD8^{+}$ T cells from wild-type or CA/+ tumour-bearing mice. **e**, Tumour growth curve of wild-type, $CD8^{-/-}$, CA/+, or *Foxp3^{Cre}Foxo1CA/+* $CD8^{-/-}$ (CA/+ $CD8^{-/-}$) mice that received orthotopic inoculation of AT-3 tumour cells. $n=4$. * $P<0.05$; ** $P<0.01$; **** $P<0.0001$; log-rank test (**b**) and two-way analysis of variance (ANOVA) (**e**). Data are mean \pm s.e.m.

A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter

Michael L. Oldham^{1,2}, Richard K. Hite^{1,2}, Alanna M. Steffen², Ermelinda Damko¹, Zongli Li^{2,3}, Thomas Walz¹ & Jue Chen^{1,2}

Cellular immunity against viral infection and tumour cells depends on antigen presentation by major histocompatibility complex class I (MHC I) molecules. Intracellular antigenic peptides are transported into the endoplasmic reticulum by the transporter associated with antigen processing (TAP) and then loaded onto the nascent MHC I molecules, which are exported to the cell surface and present peptides to the immune system¹. Cytotoxic T lymphocytes recognize non-self peptides and program the infected or malignant cells for apoptosis. Defects in TAP account for immunodeficiency and tumour development. To escape immune surveillance, some viruses have evolved strategies either to downregulate TAP expression or directly inhibit TAP activity. So far, neither the architecture of TAP nor the mechanism of viral inhibition has been elucidated at the structural level. Here we describe the cryo-electron microscopy structure of human TAP in complex with its inhibitor ICP47, a small protein produced by the herpes simplex virus I. Here we show that the 12 transmembrane helices and 2 cytosolic nucleotide-binding domains of the transporter adopt an inward-facing conformation with the two nucleotide-binding domains separated. The viral inhibitor ICP47 forms a long helical hairpin, which plugs the translocation pathway of TAP from the cytoplasmic side. Association of ICP47 precludes substrate binding and prevents nucleotide-binding domain closure necessary for ATP hydrolysis. This work illustrates a striking example of immune evasion by persistent viruses. By blocking viral antigens from entering the endoplasmic reticulum, herpes simplex virus is hidden from cytotoxic T lymphocytes, which may contribute to establishing a lifelong infection in the host.

Inside our body, every nucleated cell has surface ‘barcodes’ that are surveyed by the immune system. These barcodes are peptides derived from intracellular proteins, presented on the surface by MHC I molecules to indicate whether the cell is healthy (reviewed in ref. 1). Peptides generated from normal cellular proteins are ignored by cytotoxic T cells, whereas viral-derived or malignant peptides will trigger an adaptive immune response, resulting in elimination of the infected or tumour cells. The peptide repertoire is generated in the cytoplasm, mainly by the proteasome, but also in part by cytosolic peptidases (Fig. 1a). Peptide uploading onto MHC I molecules takes place inside the endoplasmic reticulum (ER) and is orchestrated by a macromolecular assembly collectively called the MHC class I peptide-loading complex (PLC). Cytosolic peptides are delivered across the ER membrane by the ATP-binding cassette (ABC) transporter TAP. The chaperones calnexin and calreticulin stabilize nascent MHC I molecules awaiting peptides. The tapasin/ERp57 heterodimer brings MHC I molecules and TAP within close proximity and catalyses peptide loading. Peptide-loaded MHC I molecules are then released from the ER and transported to the cell surface for antigen presentation.

As the MHC I antigen presentation pathway plays a crucial role in eradicating intracellular pathogens, it is not surprising that some viruses have evolved the ability to interfere with this process (reviewed in ref. 2). The peptide transporter TAP in particular is a primary target for viral evasion (reviewed in ref. 3). TAP is a heterodimeric ABC transporter that contains two subunits, TAP1 and TAP2, which share 37% sequence identity and are predicted to have similar structures. Each subunit contains an amino (N)-terminal

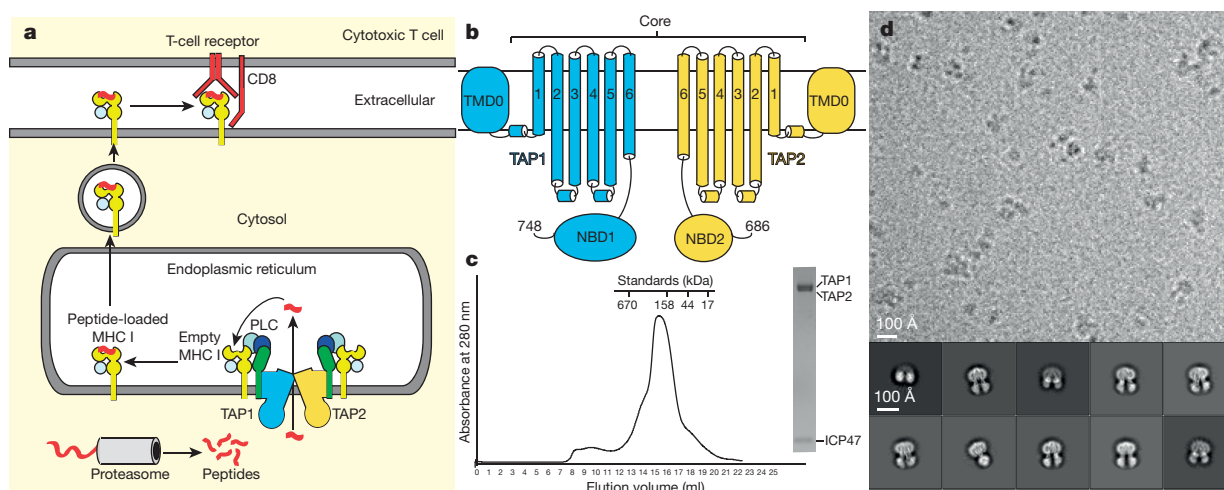


Figure 1 | Purification and cryo-EM characterization of TAP. **a**, The MHC class I antigen presentation pathway. PLC: peptide-loading complex. **b**, Topology diagram of TAP1 (blue) and TAP2 (gold). The residue numbers of the C termini are indicated. **c**, Gel-filtration profile of the

TAP-ICP47 complex. Inset: SDS-polyacrylamide gel electrophoresis of the peak fraction stained with Coomassie blue. **d**, A typical micrograph of the TAP-ICP47 complex after drift correction. Also shown are representative two-dimensional class averages of the particles.

¹The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ²Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815, USA. ³Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

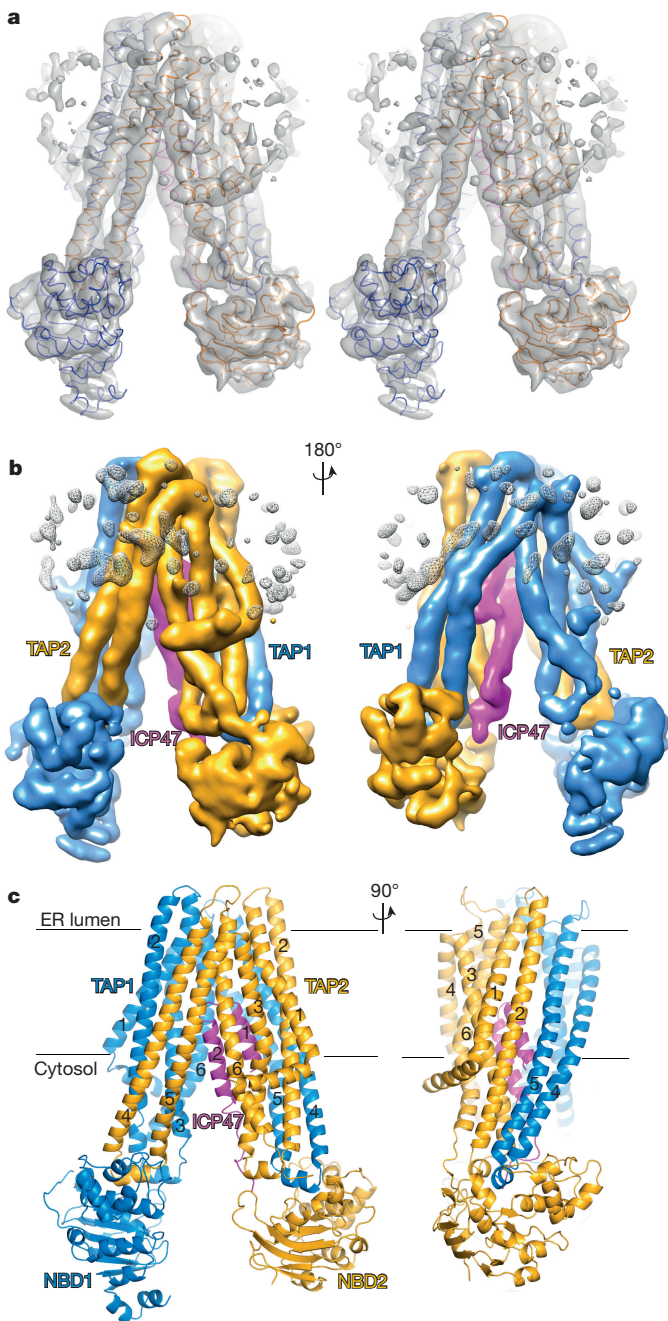


Figure 2 | Three-dimensional reconstruction. **a**, Stereo view of the overall density map, filtered to 6.5 Å. The α -carbon traces of the TAP core and the N-terminal 50 residues of ICP47 are also shown. **b**, Two views of the overall density map, coloured by protein subunit. **c**, Ribbon diagram of the poly-alanine model presented in two orthogonal views. The TM helices in the core of the transporter are labelled according to Fig. 1b. TAP1, blue; TAP2, gold; ICP47, magenta.

transmembrane region (TMD0) that interacts with tapasin, followed by six transmembrane (TM) helices that form the peptide translocation pathway and a canonical nucleotide-binding domain (NBD) that hydrolyses ATP (Fig. 1b)⁴. The core TAP, devoid of the TMD0s, is necessary and sufficient for peptide transport⁴. So far, five viral proteins have been identified as TAP inhibitors. Four are encoded by members of the herpes virus family and one by cowpox virus³. These viral inhibitors are valuable tools for selective immune suppression and for understanding the fundamental mechanism of antigen presentation.

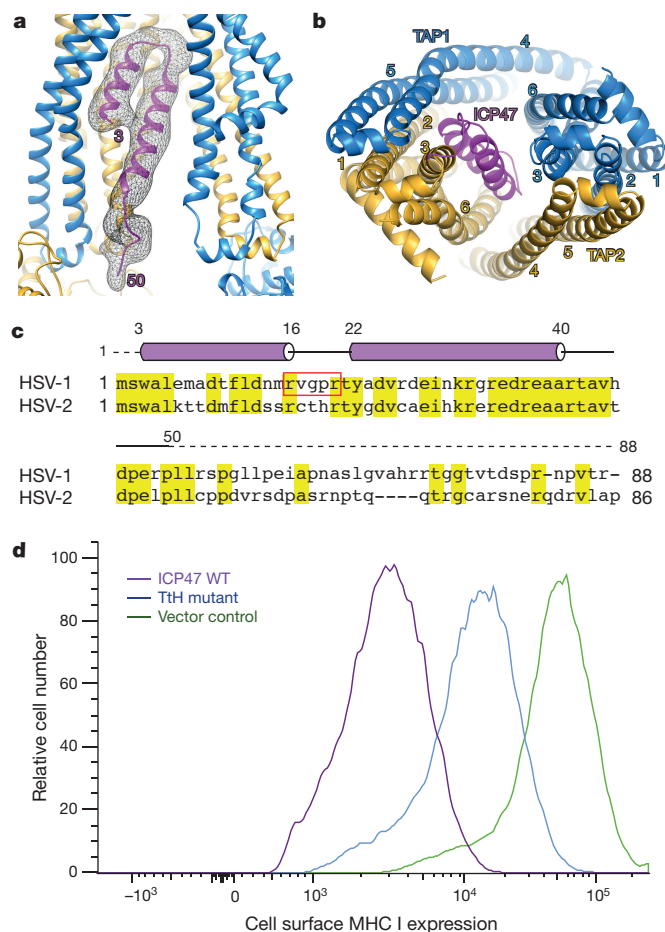


Figure 3 | The viral inhibitor ICP47 plugs into the transmembrane pathway. **a**, EM density corresponding to ICP47 (black mesh), viewed from the plane of the membrane. Terminal residues in the ICP47 model are indicated. **b**, Binding of ICP47 to TAP, viewed along the membrane normal from the cytoplasm. **c**, The helix-loop-helix structure of ICP47 is conserved between HSV-1 and HSV-2. Conserved residues are highlighted in yellow. A red box highlights residues replaced by alanine in the 'turn-to-helix' (TtH) mutant. **d**, FACS analysis (three repeats) of MHC I surface expression in cells expressing wild-type (WT) and mutant ICP47.

Here we focus our study on a TAP inhibitor encoded by herpes simplex virus (HSV). Both types of HSV, HSV-1 (oral herpes) and HSV-2 (genital herpes), somehow elude the human immune system and lead to a lifelong infection. The first clue as to how HSV bypasses the immune system came from observations that cells infected by HSV have reduced surface expression of MHC I molecules⁵ and are resistant to cytotoxic T cells⁶. Since this resistance develops within 3 hours of HSV infection, researchers narrowed their search for the responsible gene to those few expressed in the early stage of infection^{7,8}. Out of these, an 88-residue protein, ICP47, was found to bind to TAP and prevent peptide translocation into the ER^{7,8}. Consequently, empty MHC I molecules were retained in the ER and viral peptide presentation was suppressed. Subsequent studies have shown that ICP47 interacts with TAP from the cytosolic side of the membrane and somehow prevents peptide binding^{9,10}. The functional domain of ICP47 has been mapped to the N-terminal 35 residues^{11,12}, which form an extended helix-loop-helix structure in lipid bilayers¹³.

In this study, we pursued structural determination of a TAP-ICP47 complex using cryo-electron microscopy (cryo-EM). The small size of the complex (166 kDa total) and the predicted pseudo-twofold symmetry between TAP1 and TAP2 make it extremely challenging to accurately align particles for three-dimensional reconstruction. To maximize the difference between the two TAP subunits, we used

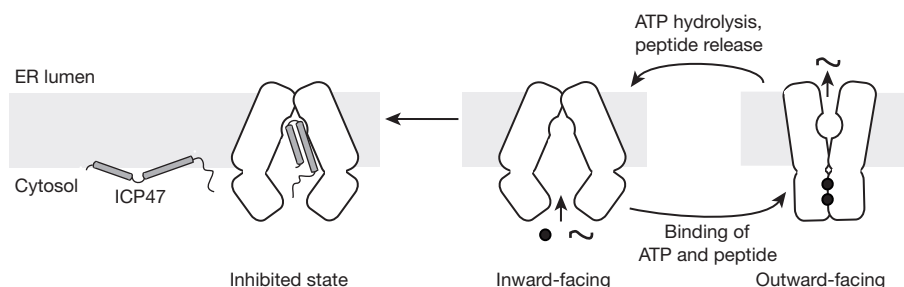


Figure 4 | ICP47 precludes peptide binding and traps TAP in an inward-facing conformation. TAP functions via alternating access, cycling between two major conformations (right). In the absence of substrates, the transporter rests in an inward-facing state in which the two NBDs are separated and the translocation pathway is exposed to the cytosol.

a shorter allele of TAP2 that lacks the last 17 amino acids in NBD2 (Fig. 1b)¹⁴. Co-expression of TAP1 and TAP2 in *Pichia pastoris* (Extended Data Fig. 1) produced a heterodimer that dissociates readily in detergents. However, by incubating the TAP-enriched membranes with ICP47 before detergent solubilization, the complex consisting of TAP1, TAP2, and ICP47 becomes more stable (Fig. 1c, d). Cryo-EM analysis of this complex (Extended Data Fig. 2 and Fig. 1d) produced a density map (Fig. 2) of an overall resolution of 6.5 Å, determined by the gold-standard refinement procedure (Extended Data Fig. 3)¹⁵. In this reconstruction the TM helices and the connectivity between the helices are clearly resolved. The density corresponding to one NBD is significantly smaller than that for the other, allowing us to confidently differentiate TAP1 from TAP2 (Fig. 2b). Most importantly, we observe strong density corresponding to the functional region of ICP47, which reveals how this viral protein inhibits peptide translocation (Fig. 2b).

The core region of TAP adopts an inverted 'V'-shaped structure, with the two TMDs making close contact on the side corresponding to the ER lumen and the NBDs separated from each other inside the cytosol (Fig. 2c). Domain swapping of TM helices 4 and 5 across the TAP1/TAP2 interface is a prominent structural feature (Fig. 2c). Indeed, the overall structure of TAP is very similar to that of other ABC exporters, including the lipid flippase MsbA from Gram-negative bacteria¹⁶, the protein transporter PCAT1 from Gram-positive bacteria¹⁷, and the multidrug transporter P-glycoprotein in eukaryotes^{18,19}. Although these transporters recognize very different substrates, they must share a common evolutionary origin and a common mechanism for coupling ATP hydrolysis to substrate translocation.

No density was observed for the N-terminal TMD0 domain of both TAP1 and TAP2 subunits (Fig. 2). Studies have shown that the TMD0s are essential to the assembly of the large peptide-loading complex^{4,20} but dispensable in peptide translocation⁴. ICP47 inhibits both full-length TAP and the core construct⁴. Our results indicate that in the absence of tapasin the two TMD0s are flexibly tethered to the core region of TAP.

Biochemical data and homology modelling suggest that the peptide translocation pathway lies at the interface of the two TMDs^{21,22}. Inside this pathway, we observe strong density consistent with the helix-loop-helix structure of ICP47 (Figs 2b and 3a). Guided by the NMR structure and secondary structure prediction¹³, we built residues 3–16 into the shorter helical density and residues 22–40 into the longer density (Fig. 3). Additional density is packed along the cytosolic region of TAP2, into which we modelled residues 41–50. The carboxy (C)-terminal region, neither required for TAP inhibition nor conserved between HSV-1 and HSV-2 (Fig. 3c), is not resolved in the EM map, suggesting high mobility.

On the basis of this model, the N-terminal half of ICP47 forms a hairpin-like structure pinned against the inner surface of TAP2 TM helices 2, 3, 6 and TAP1 TM helix 4 (Fig. 3b). The two helices of ICP47 run anti-parallel to each other, connected by a sharp turn at the top of the TM cavity. The extensive packing between TAP and ICP47 is

Upon association of substrates and ATP, the transporter undergoes a conformational change that reorients the TMDs and positions ATP at a closed NBD dimer interface for hydrolysis. ATP hydrolysis releases the substrate and resets the transporter to the resting state. ICP47 binds TAP stabilizing the inward-facing conformation (left).

consistent with the nanomolar affinity of ICP47, orders of magnitude higher than those of the substrate peptides^{9,12}.

The potency of ICP47 appears to come from its helical hairpin structure, which provides a greater interface with TAP than a typical substrate. To test whether the flexibility of the connecting loop is important, we constructed a 'turn-to-helix' mutant by replacing residues 16–22 with alanine, which has the highest propensity to form an α -helix and thus would oppose (but not preclude) the formation of a turn. ICP47 activity was measured in human epithelial cells, in which cytosolic expression of ICP47 inhibits endogenous TAP and reduces the amount of MHC I molecules expressed on the cell surface. A green fluorescent protein (GFP) tag was fused to the C terminus of ICP47 as a marker to select for cells expressing similar amounts of ICP47. Consistent with the EM structure and previous mutagenesis data¹¹, the 'turn-to-helix' mutant is much less potent than the wild-type construct. Specifically, mutant ICP47 reduced surface MHC I expression by only fivefold as opposed to a 20-fold reduction by wild-type ICP47 (Fig. 3d).

Although ICP47 competes for the same binding site, we do not believe it mimics the substrate binding process. Unlike substrates, ICP47 inhibits rather than stimulates ATP hydrolysis²³. Furthermore, in contrast to ICP47, which separates the two NBDs, substrate binding induces partial closure of the NBDs²⁴. More recently, electron paramagnetic resonance studies showed that TAP binds its substrates in their extended conformations, comparable to how MHC I molecules present peptides²⁵.

Comparison of our structure with the NMR structure of ICP47 (ref. 13) suggests that it undergoes major conformational changes upon association with TAP (Fig. 4). In isolation, the N-terminal two helices of ICP47 are flexibly linked and bind to the surface of the membrane at a slight tilt¹³. In the complex with TAP, ICP47 forms a straight hairpin and inserts perpendicularly into the membrane. Although exact determinations of the amino acid register cannot be made at the current resolution, the overall structure readily explains how ICP47 inhibits peptide transport into the ER. By plugging a long helical hairpin into the translocation pathway, ICP47 directly blocks substrates from binding. Furthermore, because ICP47 is too large to be transported by TAP, its high-affinity binding traps TAP in an inactive conformation. Like other ABC transporters, TAP functions by alternating between two major conformations, each exposing the translocation pathway to one side of the membrane (Fig. 4). Binding of ICP47 stabilizes the inward-facing conformation, and thus prevents TAP from transitioning to an outward-facing state in which the NBDs form a closed dimer and the translocation pathway orientates towards the ER lumen (Fig. 4).

In addition to viral inhibition discussed in this study, the structural basis of two cellular regulatory mechanisms has been elucidated^{26–28}, both are relevant to nutrient uptake in bacteria. As a classic example of carbon catabolite repression, when a preferred carbon source is available, bacteria suppress the uptake of maltose through direct binding of a regulatory protein to the maltose transporter²⁶. Methionine

and molybdate transporters offer another example^{27,28}. In both cases, at high intracellular concentration, the substrate binds and inhibits the corresponding transporter, a feedback mechanism that limits the amount of import into the cell^{27,28}. Unlike viral inhibition, both cellular inhibitions are allosteric and reversible, regulated by the metabolic state of the cells. One common theme among all these inhibition mechanisms is that the inhibitor binds and stabilizes the transporter in the inward-facing state, a conformation unable to hydrolyse ATP. We speculate that this strategy may be advantageous in preserving cellular energy sources. It is also possible that for most ABC transporters, the inward-facing state is most common and thus naturally targeted by regulators.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 October; accepted 1 December 2015.

Published online 20 January 2016.

- Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of antigen processing. *Annu. Rev. Immunol.* **31**, 443–473 (2013).
- van de Weijer, M. L., Luteijn, R. D. & Wiertz, E. J. Viral immune evasion: lessons in MHC class I antigen presentation. *Semin. Immunol.* **27**, 125–137 (2015).
- Verweij, M. C. *et al.* Viral inhibition of the transporter associated with antigen processing (TAP): a striking example of functional convergent evolution. *PLoS Pathog.* **11**, e1004743 (2015).
- Koch, J., Guntrum, R., Heintke, S., Kyritsis, C. & Tampé, R. Functional dissection of the transmembrane domains of the transporter associated with antigen processing (TAP). *J. Biol. Chem.* **279**, 10142–10147 (2004).
- Hill, A. B., Barnett, B. C., McMichael, A. J. & McGeoch, D. J. HLA class I molecules are not transported to the cell surface in cells infected with herpes simplex virus types 1 and 2. *J. Immunol.* **152**, 2736–2741 (1994).
- York, I. A. *et al.* A cytosolic herpes simplex virus protein inhibits antigen presentation to CD8⁺ T lymphocytes. *Cell* **77**, 525–535 (1994).
- Hill, A. *et al.* Herpes simplex virus turns off the TAP to evade host immunity. *Nature* **375**, 411–415 (1995).
- Früh, K. *et al.* A viral inhibitor of peptide transporters for antigen presentation. *Nature* **375**, 415–418 (1995).
- Tomazin, R. *et al.* Stable binding of the herpes simplex virus ICP47 protein to the peptide binding site of TAP. *EMBO J.* **15**, 3256–3266 (1996).
- Ahn, K. *et al.* Molecular mechanism and species specificity of TAP inhibition by herpes simplex virus ICP47. *EMBO J.* **15**, 3247–3255 (1996).
- Galocha, B. *et al.* The active site of ICP47, a herpes simplex virus-encoded inhibitor of the major histocompatibility complex (MHC)-encoded peptide transporter associated with antigen processing (TAP), maps to the NH₂-terminal 35 residues. *J. Exp. Med.* **185**, 1565–1572 (1997).
- Neumann, L., Kraas, W., Uebel, S., Jung, G. & Tampé, R. The active domain of the herpes simplex virus protein ICP47: a potent inhibitor of the transporter associated with antigen processing. *J. Mol. Biol.* **272**, 484–492 (1997).
- Aisenbrey, C. *et al.* Structure and dynamics of membrane-associated ICP47, a viral inhibitor of the MHC I antigen-processing machinery. *J. Biol. Chem.* **281**, 30365–30372 (2006).
- Powis, S. H. *et al.* Polymorphism in a second ABC transporter gene located within the class II region of the human major histocompatibility complex. *Proc. Natl Acad. Sci. USA* **89**, 1463–1467 (1992).
- Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).
- Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. Flexibility in the ABC transporter MsbA: alternating access with a twist. *Proc. Natl Acad. Sci. USA* **104**, 19005–19010 (2007).
- Lin, D. Y., Huang, S. & Chen, J. Crystal structures of a polypeptide processing and secretion transporter. *Nature* **523**, 425–430 (2015).
- Aller, S. G. *et al.* Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **323**, 1718–1722 (2009).
- Jin, M. S., Oldham, M. L., Zhang, Q. & Chen, J. Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. *Nature* **490**, 566–569 (2012).
- Leonhardt, R. M., Keusekotten, K., Bekpen, C. & Knittler, M. R. Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. *J. Immunol.* **175**, 5104–5114 (2005).
- Nijenhuis, M. & Hämmerling, G. J. Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site. *J. Immunol.* **157**, 5467–5477 (1996).
- Corradi, V., Singh, G. & Tieleman, D. P. The human transporter associated with antigen processing: molecular models to describe peptide binding competent states. *J. Biol. Chem.* **287**, 28099–28111 (2012).
- Gorbulev, S., Abele, R. & Tampé, R. Allosteric crosstalk between peptide-binding, transport, and ATP hydrolysis of the ABC transporter TAP. *Proc. Natl Acad. Sci. USA* **98**, 3732–3737 (2001).
- Geng, J., Sivaramakrishnan, S. & Raghavan, M. Analyses of conformational states of the transporter associated with antigen processing (TAP) protein in a native cellular membrane environment. *J. Biol. Chem.* **288**, 37039–37047 (2013).
- Herget, M. *et al.* Conformation of peptides bound to the transporter associated with antigen processing (TAP). *Proc. Natl Acad. Sci. USA* **108**, 1349–1354 (2011).
- Chen, S., Oldham, M. L., Davidson, A. L. & Chen, J. Carbon catabolite repression of the maltose transporter revealed by X-ray crystallography. *Nature* **499**, 364–368 (2013).
- Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A. & Rees, D. C. The high-affinity *E. coli* methionine ABC transporter: structure and allosteric regulation. *Science* **321**, 250–253 (2008).
- Gerber, S., Cornellas-Bigler, M., Goetz, B. A. & Locher, K. P. Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science* **321**, 246–250 (2008).

Acknowledgements We thank Z. Yu, C. Hong and R. Huang for assistance in data collection and processing. We thank X. Zhang for training in preparation of cryo-EM grids. We also thank S. McCarry for editing the manuscript. R.K.H. is a Howard Hughes Medical Institute fellow of the Helen Hay Whitney Foundation and J.C. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions M.L.O. performed protein purification and cryo-EM experiments. R.K.H. provided guidance in data processing. A.M.S. provided assistance with protein expression and purification. E.D. performed the fluorescence-activated cell sorting (FACS) experiments. Z.L. and T.W. collected preliminary cryo-EM data and generated the initial model. M.L.O. and J.C. prepared the manuscript with input from all co-authors.

Author Information The three-dimensional cryo-EM density map has been deposited in the Electron Microscopy Data Bank under the accession number EMD-6533. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C. (juechen@rockefeller.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Expression and purification of ICP47. The gene encoding human herpes simplex virus 1 (HSV-1) ICP47 was synthesized (BioBasic) and subcloned into the ligation-independent cloning vector pMCSG20 (Amp^R), which contained an N-terminal glutathione S-transferase (GST) affinity tag preceding a tobacco etch virus (TEV) protease cleavage site. BL21(DE3)RIL *Escherichia coli* cells containing the pMCSG20-ICP47 vector were grown to mid-exponential phase at 37 °C in Lysogeny broth (LB) medium and expression was induced with 200 μ M isopropyl β -D-thiogalactoside (IPTG) at 20 °C for 24 h. The cells were harvested via centrifugation (4,000g for 12 min at 4 °C) and broken by two passes through a high-pressure homogenizer (EmulsiFlex-C3; Avestin). Cell lysate supernatants were loaded onto glutathione sepharose 4B resin (GE Healthcare) equilibrated with PBS buffer, pH 7.4, containing 5 mM D,L-dithiothreitol (DTT). To remove the GST tag from ICP47, the column was equilibrated into a TEV protease cleavage buffer containing 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.5 mM EDTA, and 5 mM DTT and incubated with TEV protease overnight. Untagged ICP47 was eluted and further purified by Superdex 75 gel-filtration chromatography (GE Healthcare) in TEV-cleavage buffer. Fractions containing the protein were pooled, concentrated to 3 mg ml⁻¹, and flash-frozen in liquid nitrogen.

Co-expression of TAP1/TAP2. Synthetic human *TAP1* and *TAP2* genes were codon-optimized for expression in *P. pastoris* (BioBasic) and subcloned into pPICZ-C-XE-protein A (Zeo^R) and pPICZ-C-XE vectors (Zeo^R), respectively. To ensure co-expression at a 1:1 molar ratio, we took advantage of the compatible cohesive ends of BamHI and BglII restriction sites to generate an expression cassette as outlined in Extended Data Fig. 1. The resultant vector was linearized by PmeI digestion and transformed into a HIS⁺ strain of SMD1163 by electroporation (BioRad Gene Pulser II). Transformants were selected on yeast extract peptone dextrose sorbitol (YPDS) agar containing 800 μ g ml⁻¹ zeocin. Colonies were grown in yeast extract peptone dextrose (YPD) cultures at 28 °C until they reached an absorbance at 600 nm, $A_{600\text{ nm}}$, of 4 to seed flasks containing minimal glycerol medium (MGY), 13.4% yeast nitrogen base, and 1% glycerol at a starting $A_{600\text{ nm}}$ of 0.5. The MGY cultures were grown at 28 °C for 24 h until they reached an $A_{600\text{ nm}}$ of 20, at which point they were harvested by centrifugation (1,500g for 15 min at 4 °C) and used to seed flasks containing minimal methanol medium (MMY), 13.4% yeast nitrogen base, and 0.5% methanol at a starting $A_{600\text{ nm}}$ of 10. These MMY cultures were grown at 28 °C for 24 h before harvesting (1,500g for 15 min at 4 °C). Cell pellets were fractionated and flash-frozen in liquid nitrogen.

Purification of the TAP-ICP47 complex. Cells expressing TAP1-protein A/TAP2 were lysed using a mixer mill (Retsch Mixer Mill 400) and incubated with purified ICP47 in a buffer containing 50 mM Tris-HCl, pH 8, 500 mM NaCl, 15% glycerol, DNase I, protease inhibitors, and 2 mM TCEP for 30 min. Cells were then solubilized with 1.5% *n*-dodecyl- β -D-maltoside (DDM; Anatrace) for 2 h. The solubilized fraction was isolated by centrifugation (70,000g for 40 min at 4 °C). The TAP1-protein A/TAP2/ICP47 complex was isolated using the protein A affinity tag on TAP1 by IgG Sepharose 6 Fast Flow (GE Healthcare). After extensive washing, PreScission protease (GE Healthcare) was added to the column and incubated overnight to remove the protein A tag. The complex was eluted with additional buffer and further purified using a Superose 6 column (GE Healthcare) equilibrated with 20 mM HEPES, pH 7.4, 150 mM NaCl, 2 mM TCEP, 1 mM DDM, and 1 mM octaethylene glycol monododecyl ether (C12E8; Anatrace). The peak fraction was used to prepare cryo-EM grids.

Initial cryo-EM imaging and generation of an initial model. Vitrified specimens of TAP-ICP47 complex were prepared on glow-discharged Quantifoil holey carbon grids by plunge-freezing into liquid ethane using a Vitrobot (FEI). Cryo-EM data were collected at liquid-nitrogen temperature using a K2 Summit direct electron detector camera (Gatan) on a Tecnai F20 electron microscope (FEI) operating at 200 keV. Dose-fractionated image stacks were recorded with UCSF Image 4 (ref. 29) in super-resolution counting mode at a calibrated magnification of $\times 40,410$ (nominal magnification of $\times 29,000$) with a dose rate of 8 electrons per pixel per second (5.2 electrons per square ångström per second). Frames were read out every 200 ms and 30 frames were collected, resulting in an exposure time of 6 s and a total dose of 31.2 electrons per square ångström. Dose-fractionated image stacks were twice binned and motion-corrected, as described²⁹. The defocus was determined with CTER³⁰. BOXER was used to interactively pick 28,813 particles from ~ 750 images³¹. The particle images were subjected to the iterative stable alignment and clustering (ISAC) procedure³² implemented in SPARX³³. Four ISAC generations specifying 100 particles per group and a pixel error threshold of 0.7 yielded 324 averages. Of these,

270 averages were used to calculate an initial three-dimensional density map with the validation of individual parameter reproducibility (VIPER) procedure in SPARX.

EM sample preparation and imaging for the final three-dimensional reconstruction. Cryo-EM grids were prepared by pipetting 3 μ l freshly purified TAP-ICP47 (2 mg ml⁻¹) onto glow-discharged C-flat holey carbon CF-1.2/1.3-4C grids (Protochips) and letting the sample adsorb for 20 s. The grids were blotted for 4 s at 90% humidity using a Vitrobot Mark IV (FEI) and immediately plunge-frozen in liquid-nitrogen-cooled liquid ethane. The grids were imaged using a FEI Titan Krios electron microscope operating at an acceleration voltage of 300 keV. Images were recorded using a K2 Summit direct electron detector (Gatan) set to super-resolution counting mode with a super-resolution pixel size of 0.675 Å using the program SerialEM³⁴. In addition, a Gatan Imaging filter with a slit width of 20 eV was used to remove inelastically scattered electrons. Movie frames were recorded with an exposure time of 200 ms using a dose rate of 10 electrons per pixel per second or 5.5 electrons per square ångström per second (1.35 Å at the image plane). Three data sets were recorded using different total doses and defocus ranges (Extended Data Table 1 and Extended Data Fig. 2).

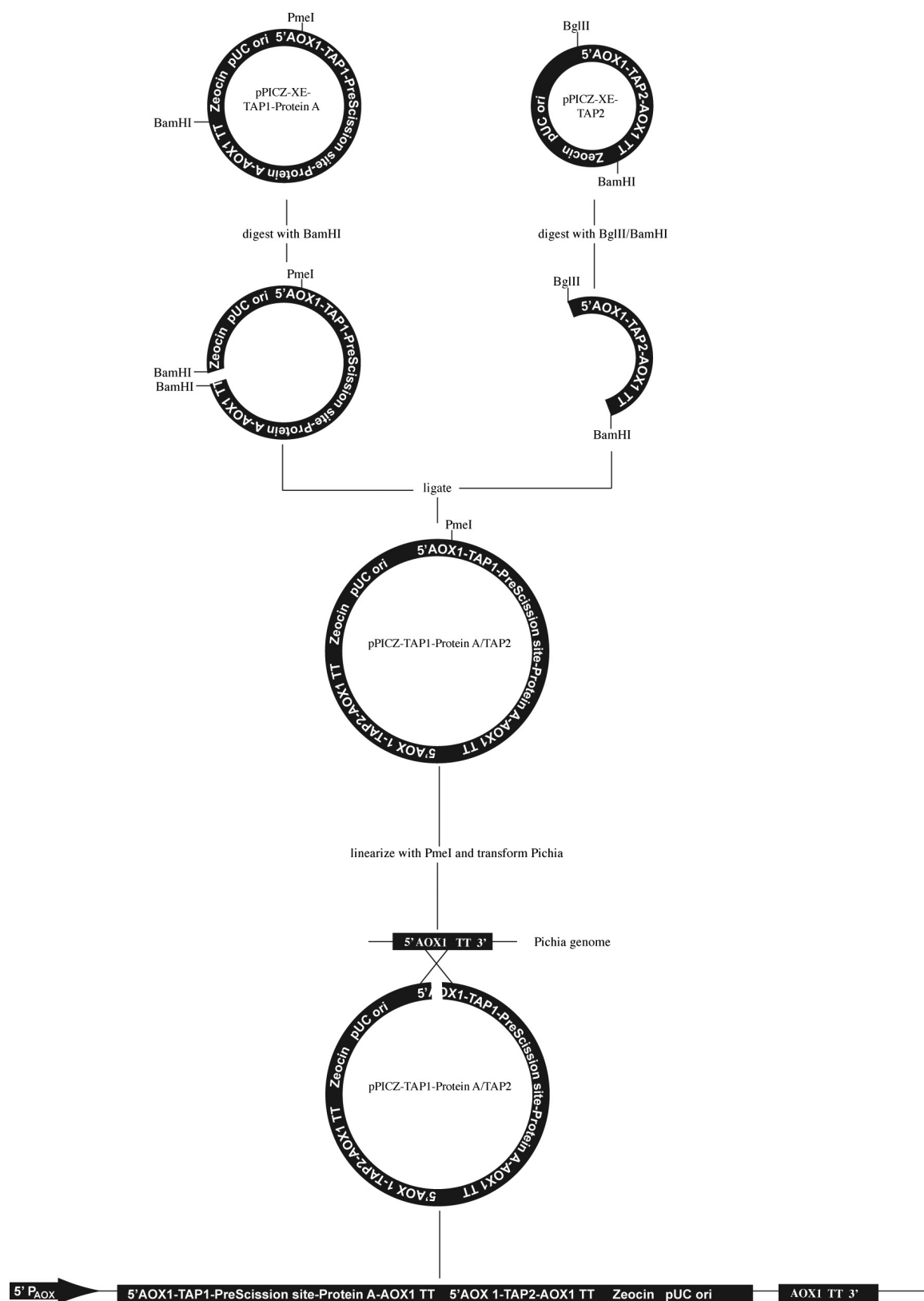
Image processing. Movie frames were corrected for gain reference and binned by a factor of 2, giving a pixel size of 1.35 Å. Drift correction was performed using the program Unblur^{35,36}. Next, the drift-corrected frames were summed into single micrographs, which were used to estimate the contrast transfer function (CTF) using CTFFIND4³⁷. The program Summovie was used to recalculate the summed images, first with a low-pass filter for autoticketing and then with the noise power restored after filtering for particle extraction³⁸. Autoticketing, particle extraction, two-dimensional classification, three-dimensional classification, and initial three-dimensional refining were all performed in Relion³⁹. To achieve a more robust classification of the extracted particles, two-dimensional classification was performed with a particle mask diameter of 145 Å while ignoring the effects of the CTF until the first zero transition. Three-dimensional classification was performed on particles from selected two-dimensional classes using the initial model calculated in SPARX as a reference map. Particles from selected three-dimensional classes from both data sets were combined for three-dimensional refinement. Using the orientation parameters determined by Relion, three-dimensional refinement in FREALIGN was also performed⁴⁰. The final map, reconstructed from 139,293 particles, had a resolution of 6.5 Å as determined by Fourier shell correlation (FSC) of independently refined half-data sets using the 0.143 cut-off criterion (Extended Data Fig. 3).

Model building. We used a model of the TAP1 NBD from a previously reported structure of the isolated domain (PDB accession number 1JJ7)⁴¹ to generate a homology model for the TAP2 NBD using the program Modeller⁴². We also generated a homology model of the TAP1 and TAP2 TMDs using the half-transporter subunit of human ABCB10 (PDB accession number 4AYT) as the source structure⁴³. We manually docked these poly-alanine models into our final cryo-EM map and rebuilt each model in Coot⁴⁴.

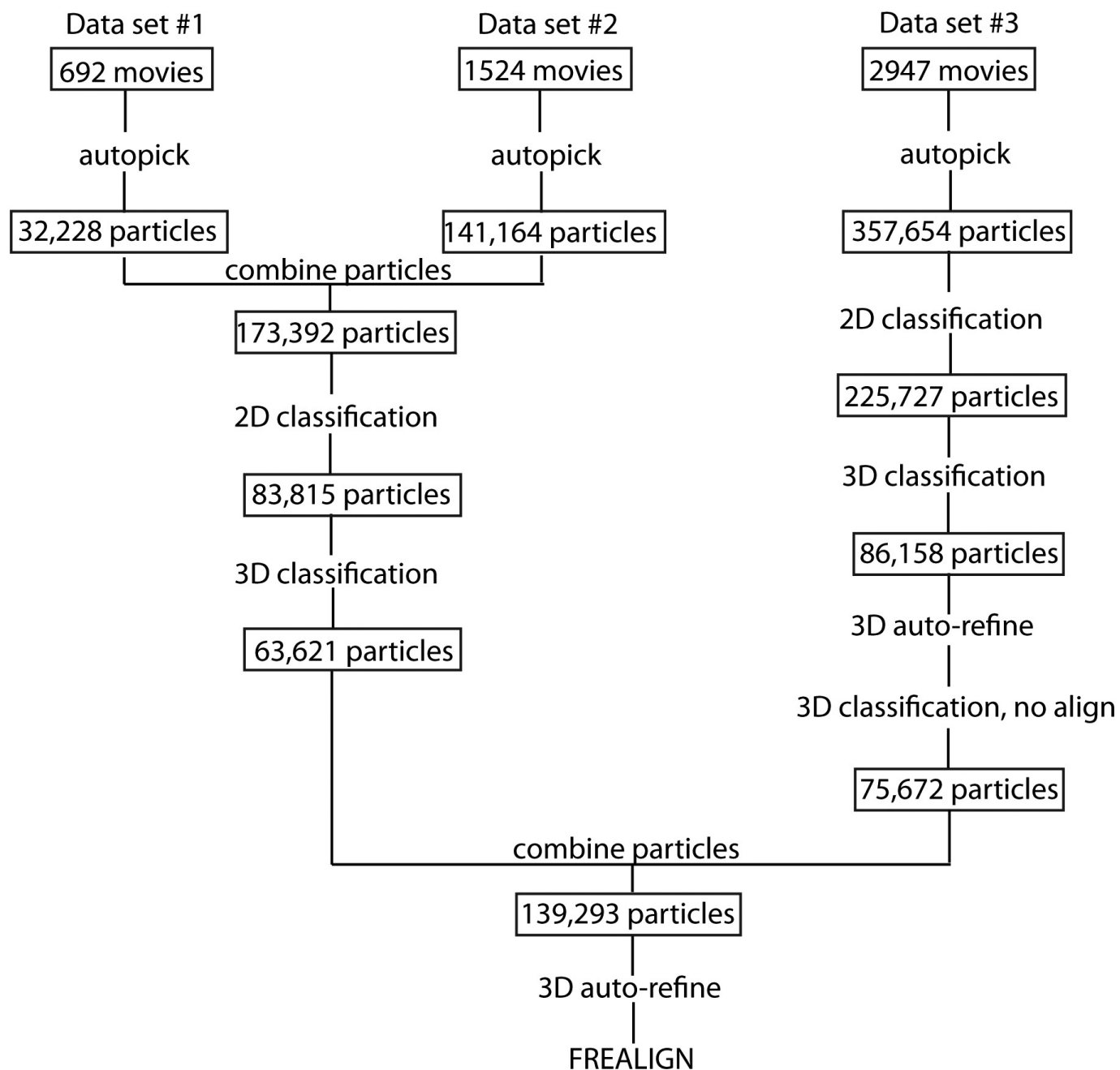
Flow cytometry analysis of MHC I surface expression. MHC I surface expression was analysed using the phycoerythrin-coupled antibody W6/32 (Abcam), which recognizes a monomorphic epitope shared among MHC class I molecules. Genes encoding HSV-1 wild-type ICP47 and the 'turn-to-helix' (TtH) mutant were cloned into a modified version of the pCDNA 3.1 vector (Life Technologies) that added a C-terminal enhanced GFP tag. HeLa cells (ATCC CCL-2) were seeded in six-well plates at a density of 5×10^5 cells per well and transfected with wild-type ICP47, the TtH mutant, or the empty vector. Cells were detached from the plate using trypsin-EDTA (0.05%) at 72 h after transfection and washed in ice-cold FACS buffer (Ca²⁺/Mg²⁺-free phosphate buffer, 10% FCS, 1% sodium azide) and centrifuged at 400g for 5 min at 4 °C. Non-specific binding was blocked by incubating the cells with phosphate buffer containing 5% (w/v) bovine serum albumin (BSA) for 15 min on ice. Antibody was added at 5 μ g ml⁻¹ and incubated for 30 min at 4 °C in the dark. Subsequently, the cells were washed three times in FACS buffer, resuspended at a density of 3×10^6 cells per millilitre, and counted using a BD LSR II Flow Cytometer (BD Biosciences). The cells were analysed at wavelengths 405 nm for DAPI nuclear stain, 488 nm for GFP fluorescence, and 561 nm for phycoerythrin fluorescence. Only live, single cells with the same levels of GFP fluorescence were used in phycoerythrin gating to compare MHC class I expression. The flow cytometry data were analysed using FlowJo 10.1 single cell analysis software (Tree Star). All experiments were repeated three times. The cell line was tested for mycoplasma contamination by PCR using a Universal Mycoplasma Detection Kit (ATCC 30-1012K).

Figure preparation. Figures were prepared using the programs PyMOL⁴⁵, Chimera⁴⁶, and FlowJo 10.1.

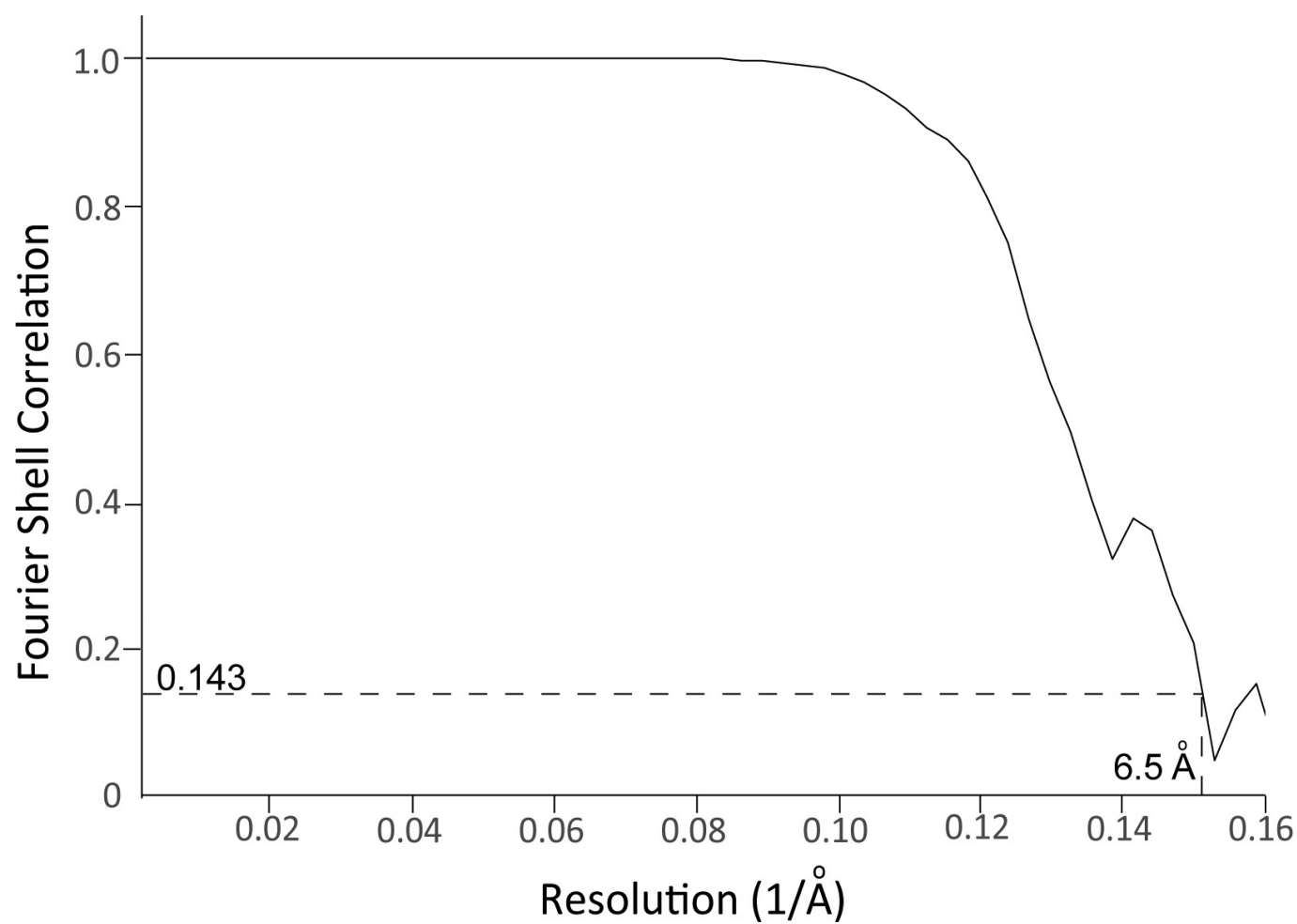
29. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
30. Penczek, P. A. *et al.* CTER-rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140**, 9–19 (2014).
31. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
32. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247 (2012).
33. Hohn, M. *et al.* SPARX, a new environment for cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
34. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
35. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **177**, 630–637 (2012).
36. Campbell, M. G. *et al.* Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**, 1823–1828 (2012).
37. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
38. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).
41. Gaudet, R. & Wiley, D. C. Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J.* **20**, 4964–4972 (2001).
42. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinform.* **Chapter 5**, Unit 5.6 (2006).
43. Shintre, C. A. *et al.* Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc. Natl Acad. Sci. USA* **110**, 9710–9715 (2013).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
45. DeLano, W. L. The PyMOL molecular graphics system (DeLano Scientific, 2002).
46. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).



Extended Data Figure 1 | Cloning strategy for TAP1-protein A/TAP2 co-expression.



Extended Data Figure 2 | Cryo-EM data processing flowchart.



Extended Data Figure 3 | FSC indicating the resolution of the density map. FSC plots were generated between reconstructions from random halves of the data. The frequency at which the dashed line passes through $FSC = 0.143$ indicates the reported resolution. Corresponding values are given in Extended Data Table 1.

Extended Data Table 1 | Summary of cryo-EM data

Data collection

Microscope	Titan Krios I, 300 keV (FEI)
Detector	K2 Summit direct electron detector (Gatan)
Pixel size	1.35 Å
Energy filter	20 eV (Gatan)

Dataset # 1

Movies	692
Frames	20
Dose	2.2 electrons/Å ² /frame
Defocus range	-1.6 to -3.0 µm

Dataset # 2

Movies	1524
Frames	40
Dose	2.2 electrons/Å ² /frame
Defocus range	-2.0 to -5.0 µm

Dataset # 3

Movies	2947
Frames	40
Dose	2.2 electrons/Å ² /frame
Defocus range	-1.5 to -3.5 µm

Final Reconstruction

Number of particles	139,293
Accuracy of rotations	2.78 degrees
Accuracy of translations	0.98 pixels
Overall resolution	6.5 Å
B-factor correction	-500 Å ²

Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53

Denise Martinez-Zapien¹, Francesc Xavier Ruiz², Juline Poirson¹, André Mitschler², Juan Ramirez¹, Anne Forster¹, Alexandra Cousido-Siah², Murielle Masson¹, Scott Vande Pol³, Alberto Podjarny², Gilles Travé¹ & Katia Zanier¹

The p53 pro-apoptotic tumour suppressor is mutated or functionally altered in most cancers. In epithelial tumours induced by 'high-risk' mucosal human papilloma viruses, including human cervical carcinoma and a growing number of head-and-neck cancers¹, p53 is degraded by the viral oncoprotein E6 (ref. 2). In this process, E6 binds to a short leucine (L)-rich LxxLL consensus sequence within the cellular ubiquitin ligase E6AP³. Subsequently, the E6/E6AP heterodimer recruits and degrades p53 (ref. 4). Neither E6 nor E6AP are separately able to recruit p53 (refs 3, 5), and the precise mode of assembly of E6, E6AP and p53 is unknown. Here we solve the crystal structure of a ternary complex comprising full-length human papilloma virus type 16 (HPV-16) E6, the LxxLL motif of E6AP and the core domain of p53. The LxxLL motif of E6AP renders the conformation of E6 competent for interaction with p53 by structuring a p53-binding cleft on E6. Mutagenesis of critical positions at the E6–p53 interface disrupts p53 degradation. The E6-binding site of p53 is distal from previously described DNA- and protein-binding surfaces of the core domain. This suggests that, in principle, E6 may avoid competition with cellular factors by targeting both free and bound p53 molecules. The E6/E6AP/p53 complex represents a prototype of viral hijacking of both the ubiquitin-mediated protein degradation pathway and the p53 tumour suppressor pathway. The present structure provides a framework for the design of inhibitory therapeutic strategies against oncogenesis mediated by human papilloma virus.

Papilloma viruses are small DNA viruses, which infect the mucosal and cutaneous epithelia of most vertebrate species. HPV-16 is the most prevalent and best studied 'high-risk' mucosal human papilloma virus (hrm-HPV), responsible for 50% of cervical carcinomas and for most HPV-positive head-and-neck cancers¹. The HPV oncoproteins E6 and E7 recognize numerous host proteins, in large part by hijacking cellular domain-motif interaction networks⁶. In particular, most mucosal and cutaneous E6 proteins recognize cellular acidic leucine (L)-rich LxxLL motifs (reviewed in ref. 7). In a recent structural study⁸, we have shown that LxxLL motifs bind to a conserved pocket of E6, which is contributed by the protein's amino (N)- and carboxy (C)-terminal zinc-binding domains (E6N and E6C) and helix linker.

In E6-mediated degradation of p53, hrm-HPV E6 proteins interact with the LxxLL motif of E6AP, leading to recruitment and polyubiquitination of p53. The isolated LxxLL peptide of E6AP (named e6ap from hereon) is sufficient to render E6 liable to interact with p53 (ref. 5). Furthermore, several studies indicate that the 'core' (DNA binding) domain of p53 is required for the interaction with E6/E6AP^{9–11}. We thus proceeded to reconstitute a minimal E6/E6AP/p53 ternary complex *in vitro* (Extended Data Fig. 1). The solubility enhanced HPV-16 E6 4C/4S mutant (named E6 from hereon), which degrades p53 with wild-type efficiency¹², was assembled with e6ap

(sequence E¹L²T³L⁴Q⁵E⁶L⁷L⁸G⁹E¹⁰E¹¹R¹²) fused to a crystallization-prone mutant of the maltose binding protein (MBP)⁸ (Extended Data Fig. 2). The resulting E6/MBP-e6ap heterodimer (named E6/e6ap from hereon) was found to interact with the isolated p53

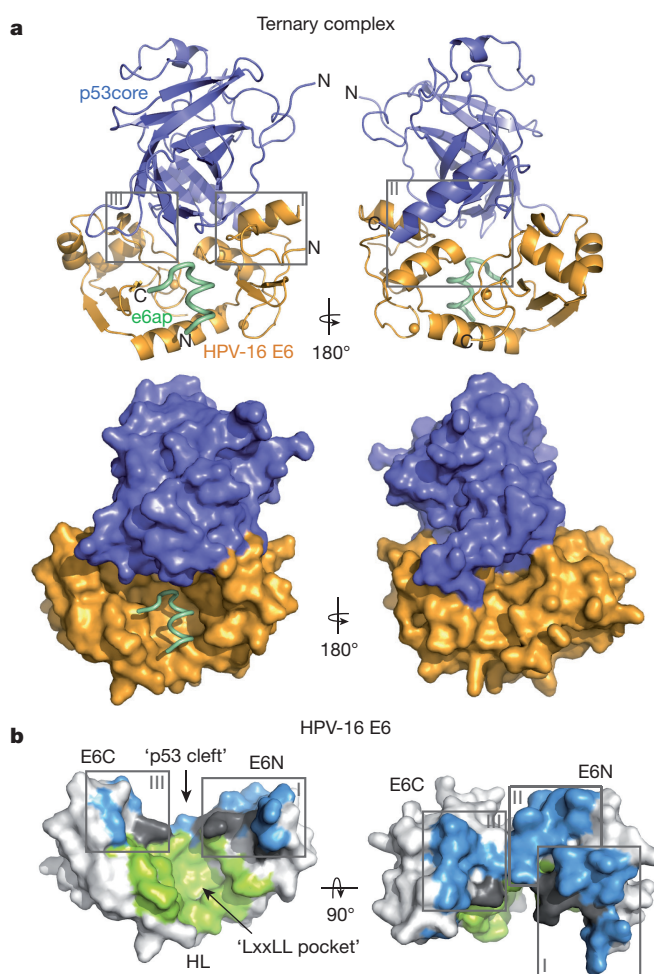


Figure 1 | Structure of the HPV-16 E6/e6ap/p53core ternary complex. **a**, Ribbon (top) and surface (bottom) representations. Green: e6ap peptide; gold: HPV-16 E6; blue: p53core. Spheres: zinc atoms. Boxes indicate sub-interfaces I–III (expanded in Fig. 2b). **b**, Surface representation of E6 coloured for residues in atomic contact with p53 (light blue), e6ap (light green), and both p53 and e6ap (dark grey). E6N and E6C: N- and C-terminal zinc-binding domains; HL: helix linker. E6 molecules on the left side of **a** and **b** are in the same orientation.

¹Equipe labellisée Ligue, Biotechnologie et signalisation cellulaire UMR 7242, Ecole Supérieure de Biotechnologie de Strasbourg, Boulevard Sébastien Brant, BP 10413, F-67412 Illkirch, France.

²Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/INSERM U964/CNRS UMR 7104/Université de Strasbourg, 1 rue Laurent Fries, BP 10142, F-67404 Illkirch, France.

³Department of Pathology, University of Virginia, PO Box 800904, Charlottesville, Virginia 22908-0904, USA.

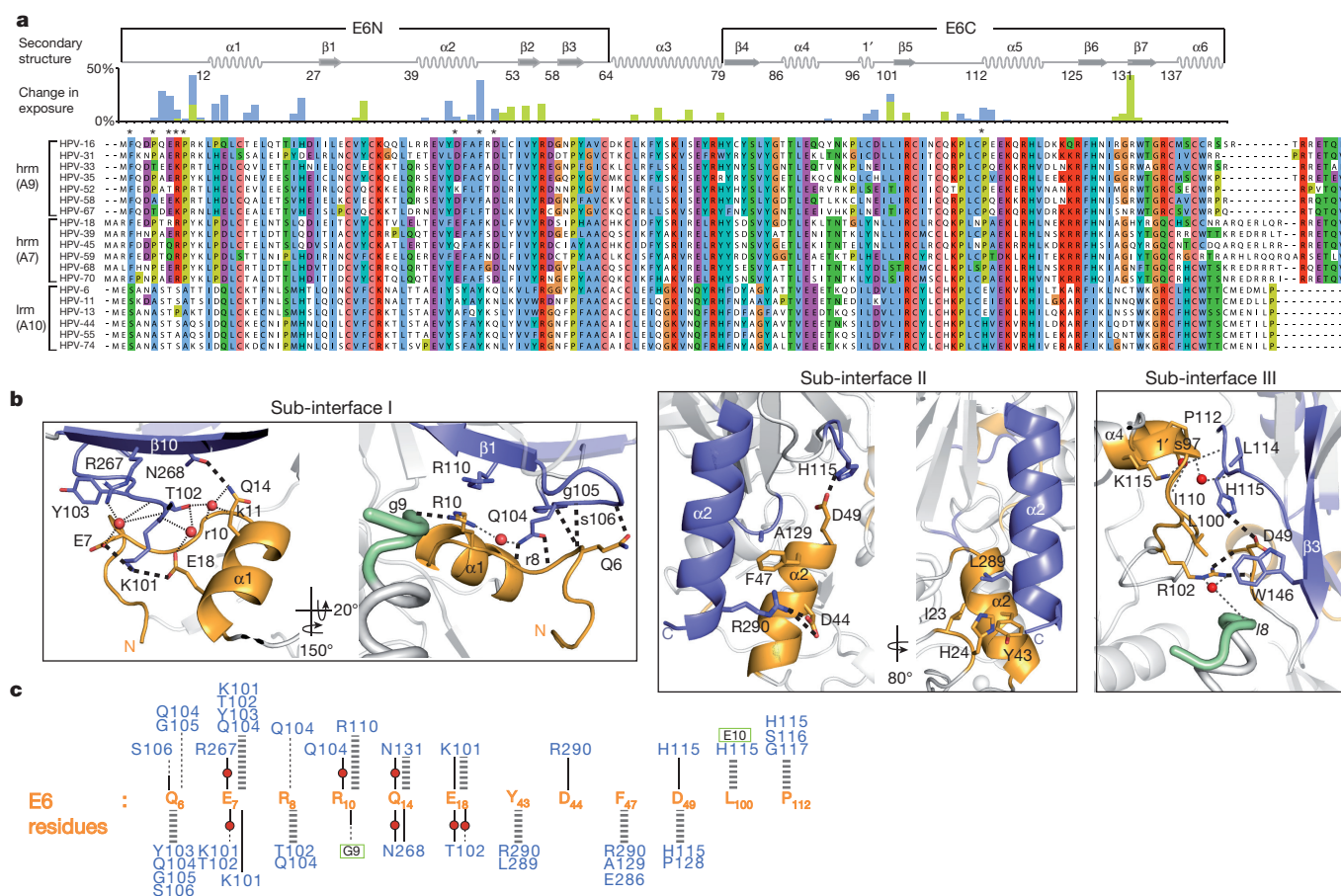


Figure 2 | Inter-molecular contacts at the E6–p53 interface. **a**, Alignment of HPV E6 sequences from hrp- and lrm-HPV groups. Histograms: burial of residues at the interface with e6ap (light green) and p53 (light blue). Asterisks indicate positions conserved in hrp-HPV. **b**, Views of sub-interfaces I–III. Red spheres: water molecules; thick dashed lines: direct polar interactions; thin dotted lines: water-mediated interactions.

core domain (residues 94–292, named p53core from hereon) by gel filtration chromatography and isothermal titration calorimetry (dissociation constant $K_d = 22 \mu\text{M}$) (Extended Data Fig. 3 and Extended Data Table 1). *In vivo*, this affinity of p53 for E6/E6AP is likely to be enhanced by avidity effects, since p53 is tetrameric and E6AP can form trimers¹³.

The E6/e6ap/p53core ternary complex raised several crystals diffracting up to 2.25 Å resolution using synchrotron radiation. This allowed structure determination by molecular replacement (Fig. 1a and Extended Data Table 2). The asymmetric unit of the crystal comprises two E6/e6ap/p53core heterotrimers, which contact each other mostly via MBP and display nearly identical structures except for the relative orientation of the MBP moieties (Extended Data Fig. 4). The structures of p53core and E6/e6ap observed in the heterotrimers are superimposable with previous structures of p53core and of E6/e6ap heterodimer, except for residues 1–8 of E6 and 10–12 of e6ap, which change conformation upon p53 binding (Extended Data Fig. 5). The similarities between the structures of the two heterotrimers in the crystal and previously solved structures of separate elements suggest that MBP does not significantly alter the overall conformation of the E6/e6ap/p53core complex.

In each heterotrimer p53core binds to a cleft, which is formed by the E6N and E6C domains and held in place by contacts tethering the domains to the e6ap peptide (Figs 1b and 2a). The E6–p53 interface covers approximately 1,200 Å². The C terminus of the e6ap peptide (residues 10–12, Extended Data Fig. 6a) also lies proximal to p53core (Extended Data Fig. 6b), but its structure is poorly defined, possibly

owing to an influence of the adjacent MBP tag. Nevertheless neither point mutations at residues 10–12 of e6ap nor extension of the e6ap peptide's C-terminal boundary altered p53 binding affinity (Extended Data Fig. 6c and Extended Data Table 1), suggesting that the e6ap C terminus does not contribute significant intermolecular contacts to p53.

The E6–p53 interface can be divided into three sub-interfaces (Figs 1a and 2b). Sub-interface I is dominated by polar interactions and brings together residues in the N-terminal arm and α1 helix of E6 and residues of the N-terminal arm, β1 and β10 strands of p53core (Fig. 2b, c). E6 residues Glu7 and Glu18 establish a bi-dentate salt bridge with Lys101 of p53 as well as water-mediated interactions with other p53 residues. In particular, Glu18 contributes to a network involving side chains and backbone groups of Gln14, Arg10 and Lys11 of E6, and of Thr102 and Asn268 of p53. Consistently, E6 mutations E18A and E18R impair ternary complex assembly (Fig. 3a) and p53 degradation (Fig. 3b). Furthermore, Gln104 and Gly105 of p53 hydrogen bond to the backbone of E6 residues Arg8 and Gln6, respectively, thus altering the conformation of the N-terminal arm of E6 (residues 1–8) (Extended Data Fig. 5).

Sub-interface II mainly involves the α2 helix of the E6N domain and the α2 helix of p53core, interacting through hydrophobic, charged and polar contacts (Fig. 2b, c). E6 residue Phe47 intercalates between p53 residues Ala129 and Arg290. In turn, Arg290 establishes a salt bridge with Asp44 of E6. These key interactions explain the reported dominant-negative phenotype of the E6 F47R mutant, which is defective for p53 degradation, restores high p53 levels and drives senescence¹⁴.

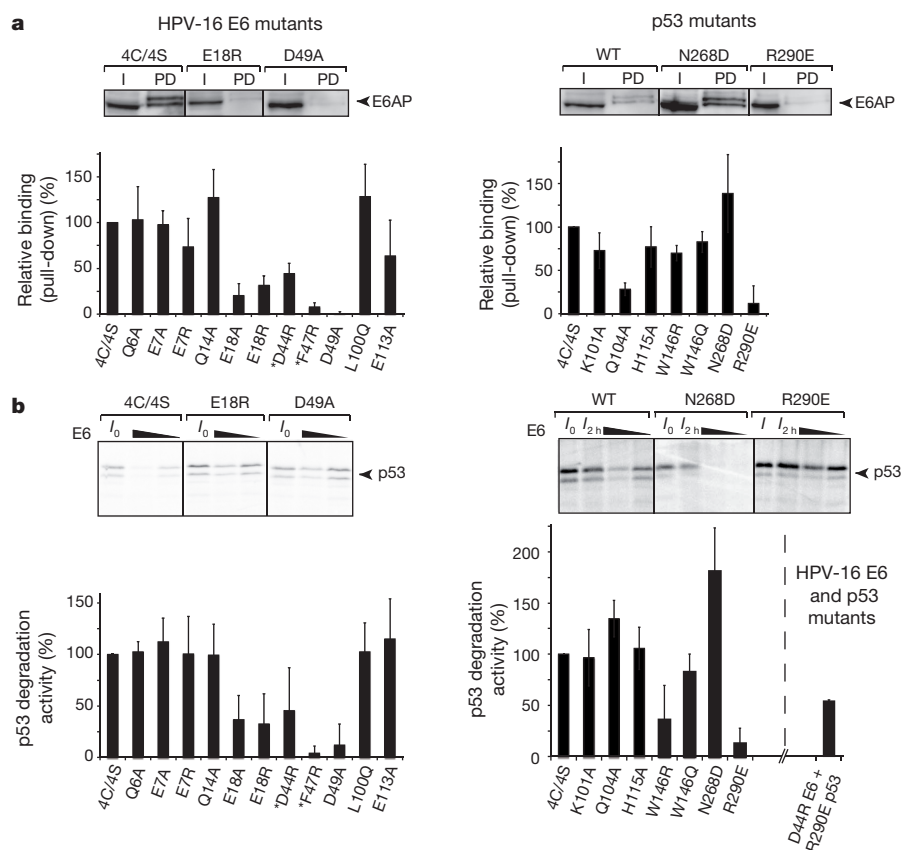


Figure 3 | Ternary complex assembly and p53 degradation activities of E6 and p53 mutants.

a, Triple pull-downs. E6AP, E6 and biotinylated p53 were co-expressed in HEK 293T cells. E6AP bound to p53 was detected by western blotting. All E6 mutations were introduced in the E6 4C/4S construct¹². Asterisks indicate previously investigated mutants¹². Top: instances of binding efficiencies. I: input; PD: pull-down. Bottom: percentage of E6AP bound to p53. See also Extended Data Fig. 8. **b**, Assays for degradation of p53. Top: instances of degradation efficiencies showing p53 before and after incubation with 5 or 2.5 μ l of E6 translation product. I₀: p53 band at time = 0; I_{2h}: p53 band at time = 2 h without E6 (see Methods). Bottom: p53 degradation activities with 2.5 μ l of E6 product. The data (mean \pm s.d.) were obtained from three independent experiments and normalized to 100% for the reference proteins (E6 4C/4S/WT p53). For gel source data see Supplementary Fig. 1.

Consistently, here we find that E6 F47R is defective for interaction with both p53core (Extended Data Fig. 3a) and full-length p53 (Fig. 3a). Besides Phe47, other E6 residues in sub-interface II (Ile23, His24 and Tyr43) provide hydrophobic contacts, whereas E6 residue Asp49 establishes polar interactions with His115 in the L1 loop of p53. Mutations disrupting the Asp44–Arg290 and Asp49–His115 interactions (D44R and D49A in E6 and the R290E in p53) impair both p53 ternary complex assembly (Fig. 3a) and p53 degradation (Fig. 3b). By contrast, creation of a swapped Glu290–Arg44 salt bridge, by combining D44R E6 and R290E p53 mutants, partly restores p53 degradation (Fig. 3b).

Finally, sub-interface III encompasses hydrophobic interactions between E6C residues Leu100 and Pro112 and p53core surface residues Leu114 and Trp146, as well as several water-mediated backbone-to-backbone contacts (Fig. 2b, c). Notably, Leu100 and Pro112 of E6C are proximal to conserved Arg102, which shapes the p53 binding cleft by mediating crucial interactions with the E6N domain and the e6ap peptide⁸.

Degradation of p53 is a hallmark activity of hrn-HPV E6 proteins^{15–17}. The structure explains the critical contributions towards p53 degradation of particular residues, which are differentially conserved in high- versus low-risk mucosal (Irm) HPV E6. In sub-interface I, most conserved positions in hrn-HPV E6 proteins correspond to Phe2, Pro5, Glu7, Arg8 and Pro9 (Fig. 2a). While Glu7 mediates direct contacts to p53, Phe2, Pro5, Arg8 and Pro9 play an indirect role at the interface by shaping the conformation of the N-terminal region (Extended Data Fig. 7). Indeed, site-directed mutagenesis of Phe2, Pro5, Arg8 or Pro9 (refs 18–20) impairs ternary complex assembly and p53 degradation activities. In sub-interface II, most conserved positions in hrn-HPV correspond to Phe47, Asp44 and Asp49 (Fig. 2a), which are here found to mediate crucial contacts to p53. These residues were formerly found to be important for p53 degradation^{14,18,21}. Notably, Phe47 and Asp44 also participate in the dimerization of the E6N domain and subsequent *in vitro* self-association of the entire E6 protein¹².

Therefore, E6 self-association and E6 binding to p53 are two distinct and competing processes mediated by partly overlapping interaction surfaces.

The LxxLL peptide of E6AP is absolutely required for E6 binding to p53, yet it does not contribute significant contacts to p53. Furthermore, in the absence of the LxxLL motif of E6AP, the few interactions connecting E6N, linker helix and E6C should be insufficient to maintain the overall E6 architecture observed in the E6/e6ap heterodimer structure (see ref. 8 for further discussion). These observations indicate that the LxxLL motif structures the p53 binding cleft on E6, thereby rendering E6 competent for interaction with p53. Interestingly, we have shown that an *in vitro* selected peptide, targeting the LxxLL pocket of HPV-16 E6, induces recruitment of p53 to E6 (refs 22, 23). Consequently, cellular proteins other than E6AP, which bind to the LxxLL pocket, might also promote the E6–p53 interaction.

The p53core is both a DNA binding domain and a protein–protein interaction hub. E6 interacts with the N-terminal arm of p53core and one of the edges of the β -sandwich. Previous mutagenesis studies had already suggested a role for these regions in E6 binding^{9,11,24}. The E6-binding interface on p53core is distal from both the DNA-binding region and protein-binding surfaces observed in all other solved complexes of p53core (Fig. 4a). Indeed, previously described protein-binding interfaces of p53core all overlap, albeit to different extents, with the DNA-binding region. In particular, the large T antigen (LTag) of oncovirus SV40, which does not degrade p53, buries the entire DNA binding interface of p53core, thereby inhibiting p53 trans-activation activity²⁵. Consistent with these observations, two frequent cancer-associated mutations (R273C and R273L), affecting a prominent DNA-binding arginine residue of p53core, abolish binding to SV40 LTag, but do not alter the interaction with E6/E6AP²⁶. Therefore, E6 might avoid competition with cellular factors by targeting both ‘free’ p53 as well as p53 bound to DNA or other proteins (Fig. 4b). This, along with the irreversible character of the degradation process, renders E6 a potent inactivator of p53 functions.

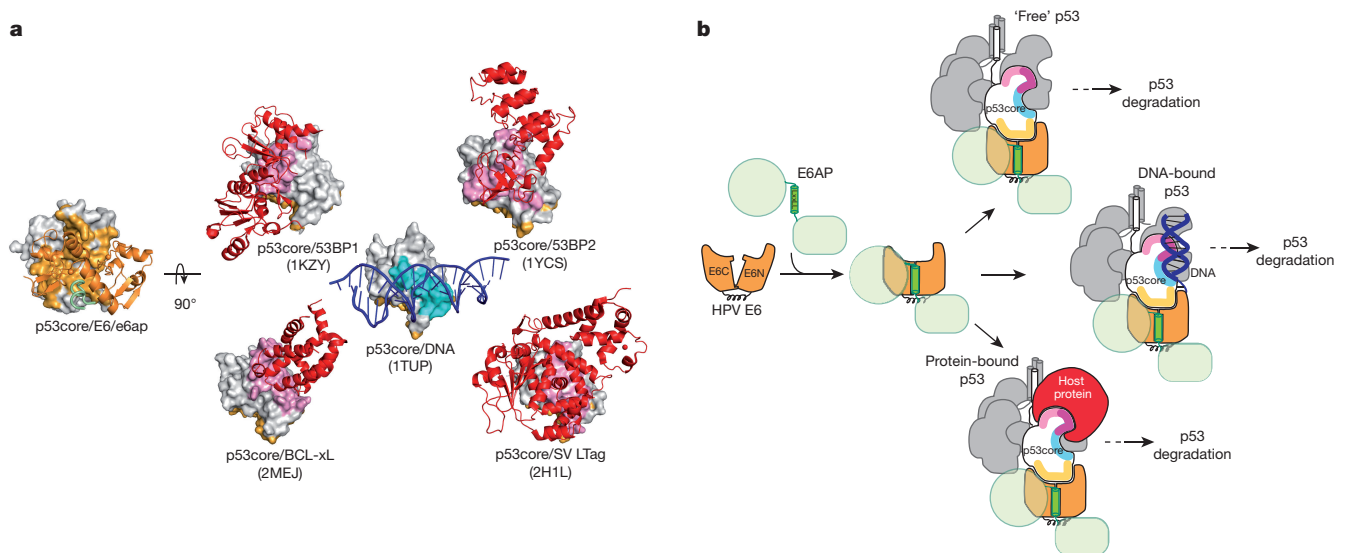


Figure 4 | p53core targeting by HPV E6. a, Complexes of p53core (surface representation) with protein or DNA partners (ribbon representation). Left: p53core bound to HPV E6/e6ap. Right: p53core bound to 53BP2, 53BP1 and BCL-xL cellular proteins, to adenovirus SV40 LTag oncoprotein and to DNA. p53core binding interfaces: E6 (light orange), other proteins (pink) and DNA (cyan). **b,** Cartoon summarizing

Recent studies employing pro-apoptotic peptides²² as well as small molecules^{27,28} directed against the LxxLL pocket have provided experimental evidence that this pocket is druggable. The p53-binding cleft observed in the present structure may represent a second potential binding site for drugs. Combinatorial strategies targeting both the LxxLL pocket and the p53-binding cleft could result in efficient disruption of the E6/E6AP/p53 complex.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 August; accepted 23 November 2015.

Published online 20 January 2016.

1. Bosch, F. X. *et al.* Comprehensive control of human papillomavirus infections and related diseases. *Vaccine* **31** (Suppl. 7), H1–H31 (2013).
2. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129–1136 (1990).
3. Huibregtse, J. M., Scheffner, M. & Howley, P. M. Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. *Mol. Cell. Biol.* **13**, 4918–4927 (1993).
4. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83 (1995).
5. Ansari, T., Brimer, N. & Vande Pol, S. B. Peptide interactions stabilize and restructure human papillomavirus type 16 E6 to interact with p53. *J. Virol.* **86**, 11386–11391 (2012).
6. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **36**, 159–169 (2011).
7. Vande Pol, S. B. & Klingelutz, A. J. Papillomavirus E6 oncoproteins. *Virology* **445**, 115–137 (2013).
8. Zanier, K. *et al.* Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. *Science* **339**, 694–698 (2013).
9. Gu, J., Rubin, R. M. & Yuan, Z. M. A sequence element of p53 that determines its susceptibility to viral oncoprotein-targeted degradation. *Oncogene* **20**, 3519–3527 (2001).
10. Li, X. & Coffino, P. High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *J. Virol.* **70**, 4509–4516 (1996).
11. Bernard, X. *et al.* Proteasomal degradation of p53 by human papillomavirus E6 oncoprotein relies on the structural integrity of p53 core domain. *PLoS ONE* **6**, e25981 (2011).
12. Zanier, K. *et al.* Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. *Structure* **20**, 604–617 (2012).
13. Ronchi, V. P., Klein, J. M., Edwards, D. J. & Haas, A. L. The active form of E6-associated protein (E6AP)/UBE3A ubiquitin ligase is an oligomer. *J. Biol. Chem.* **289**, 1033–1048 (2014).
14. Ristiani, T., Fournane, S., Orfanoudakis, G., Travé, G. & Masson, M. A single-codon mutation converts HPV16 E6 oncoprotein into a potential tumor suppressor, which induces p53-dependent senescence of HPV-positive HeLa cervical cancer cells. *Oncogene* **28**, 762–772 (2009).
15. Fu, L. *et al.* Degradation of p53 by human Alphapapillomavirus E6 proteins shows a stronger correlation with phylogeny than oncogenicity. *PLoS ONE* **5**, e12816 (2010).
16. Mesplède, T. *et al.* p53 degradation activity, expression, and subcellular localization of E6 proteins from 29 human papillomavirus genotypes. *J. Virol.* **86**, 94–107 (2012).
17. Hiller, T., Poppelreuther, S., Stubenrauch, F. & Iftner, T. Comparative analysis of 19 genital human papillomavirus types with regard to p53 degradation, immortalization, phylogeny, and epidemiologic risk classification. *Cancer Epidemiol. Biomark. Prev.* **15**, 1262–1267 (2006).
18. Foster, S. A., Demers, G. W., Etscheid, B. G. & Galloway, D. A. The ability of human papillomavirus E6 proteins to target p53 for degradation in vivo correlates with their ability to abrogate actinomycin D-induced growth arrest. *J. Virol.* **68**, 5698–5705 (1994).
19. Cooper, B. *et al.* Requirement of E6AP and the features of human papillomavirus E6 necessary to support degradation of p53. *Virology* **306**, 87–99 (2003).
20. Liu, Y. *et al.* Multiple functions of human papillomavirus type 16 E6 contribute to the immortalization of mammary epithelial cells. *J. Virol.* **73**, 7297–7307 (1999).
21. Nakagawa, S. *et al.* Mutational analysis of human papillomavirus type 16 E6 protein: transforming function for human cells and degradation of p53 *in vitro*. *Virology* **212**, 535–542 (1995).
22. Zanier, K. *et al.* The E6AP binding pocket of the HPV16 E6 oncoprotein provides a docking site for a small inhibitory peptide unrelated to E6AP, indicating druggability of E6. *PLoS ONE* **9**, e112514 (2014).
23. Stutz, C. *et al.* Intracellular analysis of the interaction between the human papillomavirus type 16 E6 oncoprotein and inhibitory peptides. *PLoS ONE* **10**, e0132339 (2015).
24. Hengsternmann, A., Linares, L. K., Ciechanover, A., Whitaker, N. J. & Scheffner, M. Complete switch from Mdm2 to human papillomavirus E6-mediated degradation of p53 in cervical cancer cells. *Proc. Natl Acad. Sci. USA* **98**, 1218–1223 (2001).
25. Lilyestrom, W., Klein, M. G., Zhang, R., Joachimiak, A. & Chen, X. S. Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev.* **20**, 2373–2382 (2006).
26. Scheffner, M., Takahashi, T., Huibregtse, J. M., Minna, J. D. & Howley, P. M. Interaction of the human papillomavirus type 16 E6 oncoprotein with wild-type and mutant human p53 proteins. *J. Virol.* **66**, 5100–5105 (1992).
27. Cherry, J. J. *et al.* Structure based identification and characterization of flavonoids that disrupt human papillomavirus-16 E6 function. *PLoS ONE* **8**, e84506 (2013).

28. Malecka, K. A. *et al.* Identification and characterization of small molecule human papillomavirus E6 inhibitors. *ACS Chem. Biol.* **9**, 1603–1612 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work received institutional support from le Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, l'Institut National de la Santé et de la Recherche Médicale (INSERM) and Région Alsace. The work was supported by grants from Ligue contre le Cancer, National Institutes of Health (grant R01CA134737 to S.V.P.), l'Agence Nationale de la Recherche (ANR-13-JSV8-004-01), Instruct (ESFR), the French Infrastructure for Integrated Structural Biology (FRISBI) and Fondation pour La Recherche Médicale (fellowship to F.X.R.F.). We thank P. Poussin-Courmontagne, E. Ennifar, V. Olieric and B. Kieffer for advice. The authors declare that the

content is solely their responsibility and does not represent the official views of the National Institutes of Health.

Author Contributions D.M.Z., J.P., A.M., J.R.R., A.F., A.C.S. and K.Z. performed experiments; F.X.R., A.P., D.M.Z. and K.Z. performed structure determination; D.M.Z., F.X.R., J.P., A.M., A.P. and K.Z. analysed the data; D.M.Z., J.P. and K.Z. prepared figures; K.Z. and GT wrote the manuscript together with comments from all authors; M.M., S.V.P., A.P., GT and K.Z. supervised the work.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession number 4XR8. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.Z. (zanier@unistra.fr) or G.T. (gilles.trave@unistra.fr).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Preparation of protein samples. DNA constructs. Residues 403–414 of human E6AP (peptide sequence E¹L²T³L⁴Q⁵E⁶L⁷L⁸G⁹E¹⁰E¹¹R¹²) were cloned via a three-alanine linker at the C terminus of a mutant MBP used to promote crystallization⁸. The point mutations introduced in MBP (D83A, K84A, K240A, E360A, K363A and D364A) have been previously described to increase the propensity of MBP to crystallize²⁹. The E6 4C/4S mutant of HPV-16 (ref. 12) (named E6 from hereon) and the core (or DNA binding) domain of human p53 (residues 94–311) were cloned into the pETM-41 vector containing an N-terminal His₆-MBP tag followed by a TEV cleavage site.

Protein expression and purification. MBP-e6ap, His₆-MBP-tev-E6 and His₆-MBP-tev-p53core constructs were overexpressed separately in *Escherichia coli* BL21(DE3) cells at 15°C for 18 h. All constructs were purified separately by amylose affinity chromatography in buffer A (50 mM Tris pH 6.8, 400 mM NaCl, and 2 mM DTT). To remove soluble aggregates, all affinity purified samples were ultracentrifuged at 110,000 g in a swing SW41 rotor (Beckman) for 16 h at 4°C. Hence p53core and E6 samples were digested by TEV protease. In the case of p53core, after TEV digestion samples were additionally purified on a heparin column. The resulting MBP-e6ap, E6 and p53 samples were thus concentrated and loaded separately onto a Superdex 75 HiLoad 16/60 gel filtration column (GE Healthcare) equilibrated in buffer A. All purification buffers were filtered, degassed and saturated with argon.

Crystallization. The E6/e6ap/p53core complex was reconstituted by mixing MBP-e6ap and E6 and p53core samples in a 1:1:1 stoichiometric ratio in buffer B (50 mM Tris pH 6.8, 200 mM NaCl, 2 mM DTT, 5 mM maltose) and concentrated to 18 mg ml⁻¹ before crystallization. Crystallization conditions were screened using commercially available kits (Qiagen, Hampton Research, Emerald Biosystems) by the sitting-drop vapour-diffusion method in 96-well MRC 2-drop plates (SWISSCI), employing a Mosquito robot (TTP Labtech). Initial crystals were obtained and used as seeds during further optimization steps.

After optimization, crystals (125 µm × 80 µm × 80 µm) grew in sitting drops made from 400 nl of protein solution at 18 mg/ml, 370 nl of reservoir solution containing 7.5% polyethylene glycol (PEG) 20 K, 50 mM MES (pH 6.5) and 30 nl of seeds. Drops were equilibrated against 50 µl of reservoir solution at 290 K. Crystals were sequentially transferred through two cryo-solutions of reservoir solution supplemented with 15% (v/v) and 25% (v/v) of PEG 200, respectively. The crystals were flash-cooled and stored in liquid nitrogen.

Data collection and processing, and structure determination. X-ray diffraction data were collected on the X06DA beamline at the Synchrotron Swiss Light Source (Villigen, Switzerland). Data were acquired from a single cryo-cooled crystal (100 K) on a Pilatus-2M detector. The 200° data were collected up to a resolution of 2.25 Å using 0.5° rotation and 0.8 s exposure time with 20% beam attenuation for each image. The data were indexed, processed and scaled using HKL2000 (ref. 30). The crystals belonged to the monoclinic space group P₂₁ with unit cell parameters $a = 78.17$ Å, $b = 129.37$ Å, $c = 82.17$ Å and $\beta = 92.4^\circ$, with a refined crystal mosaicity of 0.28–0.34°. The asymmetric unit contained two copies of the E6/MBP-e6ap/p53core heterotrimer, with a corresponding Matthew's coefficient of 2.48 Å³ per dalton and a solvent content of 50.34%. The structure was solved by two sequential molecular replacements using Phaser³¹ and the structures of MBP, E6/e6ap complex and p53core as templates (PDB accession numbers 4GIZ and 1TUP, respectively). Crystallographic refinement involved repeated cycles of conjugate-gradient energy minimization and temperature-factor refinement, and was performed using PHENIX³² followed by iterative model building in Coot³³. All the disordered residues were fitted to the built electron density map, except from e6ap residue Arg12 in one of the two heterotrimers (corresponding to chain B residue Arg383 in the PDB file).

Data collection and refinement statistics are summarized in Extended Data Table 2. The quality of the refined models was assessed using MOLPROBITY³⁴. All molecular graphics figures were made using PyMOL³⁵. E6 residues at interface regions (Figs 1b and 2a) were identified by the observation of an increase in solvent accessibility obtained upon removal of e6ap or p53core from the PDB file of the ternary complex. Positive 'change in exposure' values (Fig. 2a) indicate E6 residues in atomic contact with or p53.

The refined model and the structure factor amplitudes have been deposited in the PDB under accession number 4XR8.

DNA constructs for triple pull-down and p53 degradation assays. HPV-16 E6 mutations were inserted in the background of the E6 4C/4S construct to prevent cross-linking of E6 molecules through intermolecular cysteine bridges. We have

previously shown that E6 4C/4S and wild-type HPV-16 E6 have indistinguishable p53 degradation activities *in vitro*¹². The p53 mutations were inserted in the background of wild-type full-length human p53. All DNA constructs were verified by sequencing.

Triple pull-downs. DNA constructs encoding for full-length HPV-16 E6 and p53, and a large fragment of E6AP (residues 291–875 of isoform II), were inserted into the final destination vectors by Gateway cloning (Invitrogen). HPV-16 E6 and E6AP were inserted into the GPCA pSPICA-N2 and pSPICA-N1 vectors respectively. GPCA vectors pSPICA-N1 and pSPICA-N2 (both derived from the pCiNeo mammalian expression vector) allow expression of test proteins as fusions to the C terminus of Gluc1 and Gluc2 complementary fragments of the *Gaussia princeps* luciferase, respectively³⁶. By contrast, p53 constructs were inserted in the BioEase-DEST vector³⁶, which incorporates a 72-amino-acid sequence from *Klebsiella pneumoniae* that directs *in vivo* biotinylation.

Degradation assays. HPV-16 E6 and full-length p53 proteins were cloned in the pXJ40 vector.

Triple pull-downs. HEK-293T cells were grown and maintained in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% FCS and 50 µg ml⁻¹ of gentamycin at 37°C with 5% CO₂ and 95% humidity. Cells were seeded in six-well plates at a concentration of 3×10^5 cells per well. After 24 h, cells were transfected using JetPEI (Polyplus transfection) with 1 µg pSPICA-N2 plasmid expressing E6, 1 µg of pSPICA-N1 plasmid expressing E6AP protein and 1 µg of BioEase plasmid expressing p53. At 24 h after transfection, cells were harvested and lysed by freeze-thawing in 100 µl of *Renilla* lysis buffer (Promega, E2820). Cellular lysates were cleared by centrifugation at 11,000 g in a microfuge for 15 min at 4°C. Subsequently, 30 µl of pre-equilibrated Streptavidin Mag Sepharose beads (GE Healthcare) were incubated with 60 µl of cellular lysate supernatant for 2 h at 4°C, thereby allowing capture of biotinylated p53. Streptavidin beads were washed three times with TNE buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP40, 1 mM EDTA and protease inhibitor), resuspended with 25 µl of protein loading buffer and loaded on a 15% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) gel. E6 and E6AP proteins were detected by western blotting using rabbit anti-Gluc antibody (NEB, reference E8023S), whereas p53 was detected using a mouse anti-p53 DO-1 antibody (Life Technologies, reference 13–4000). The immunoreactive bands were visualized using WesternBright Sirius (Advansta) and an LAS 4000 camera. Bands were quantified using ImageGauge software (Fujifilm). Error bars, s.d. from three independent experiments.

Cell lines. The HEK-293T cell line was provided by P. Charneau. These cells were authenticated and tested to be mycoplasma free.

p53 degradation assays. E6 and p53 proteins were *in vitro* translated in the presence of [³⁵S]methionine using the TNT T7 coupled rabbit reticulocyte system. The p53 degradation reactions were performed in 10 µl volumes by incubating 2 µl of p53 translation product with either 5 or 2.5 µl of E6 translation products at 28°C for 2 h according to previously described protocols². Reactions were resolved on a 15% SDS-PAGE gel. Gels were exposed to a PhosphorImager screen and scanned using a Typhoon FLA 9500 imaging system (GE Healthcare). Reactive p53 bands were quantified using ImageQuant TL software (GE Healthcare).

The p53 degradation activity values reported in Fig. 3 were derived using the formula $(I_0 - I)/I_0$, where I is the intensity of the p53 double band after incubation with E6 and I_0 the p53 signal in the input lane. The p53 degradation activity values were normalized to 100% for the reference proteins (E6 4C/4S/WT p53). For p53 mutant proteins (Fig. 3b), besides the I_0 input control, the I_{2h} control (corresponding to p53 at time = 2 h in the absence of E6) was added to check for the intrinsic stability of p53 mutants. Error bars, s.d. from three independent experiments.

29. Moon, A. F., Mueller, G. A., Zhong, X. & Pedersen, L. C. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci.* **19**, 901–913 (2010).

30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

31. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

32. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

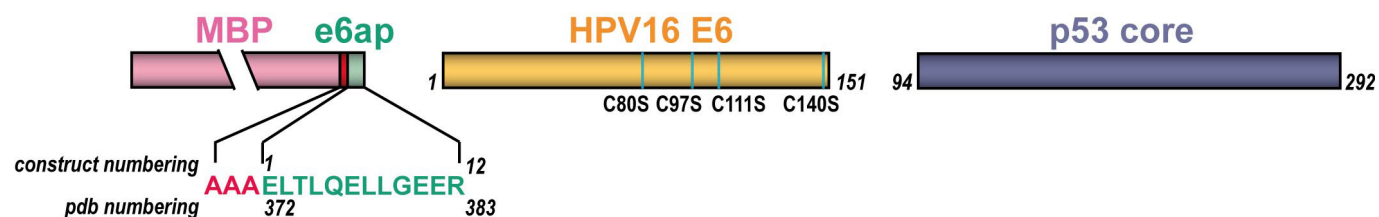
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

34. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

35. DeLano, W. L. *The PyMOL molecular graphics system* (DeLano Scientific, 2002).

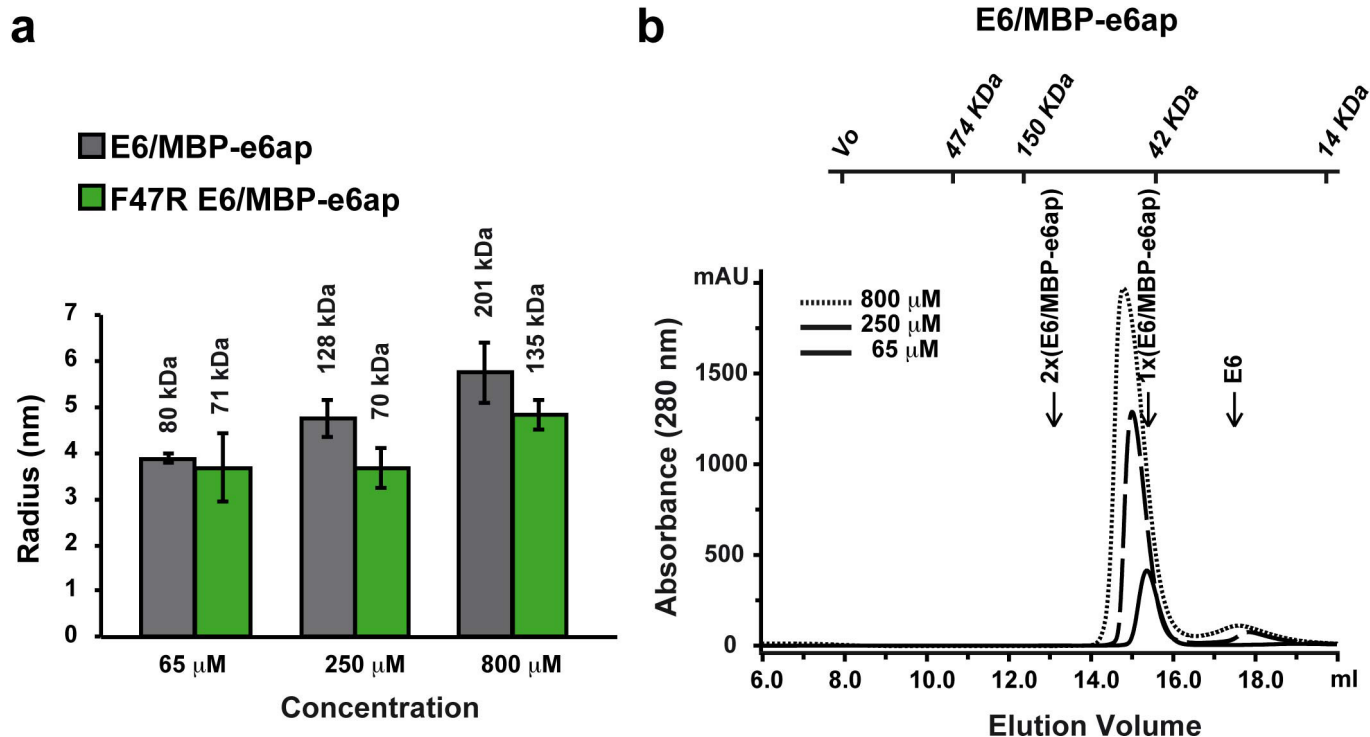
36. Cassonnet, P. *et al.* Benchmarking a luciferase complementation assay for detecting protein complexes. *Nature Methods* **8**, 990–992 (2011).

37. Wang, Y., Rosengarth, A. & Luecke, H. Structure of the human p53 core domain in the absence of DNA. *Acta Crystallogr. D* **63**, 276–281 (2007).



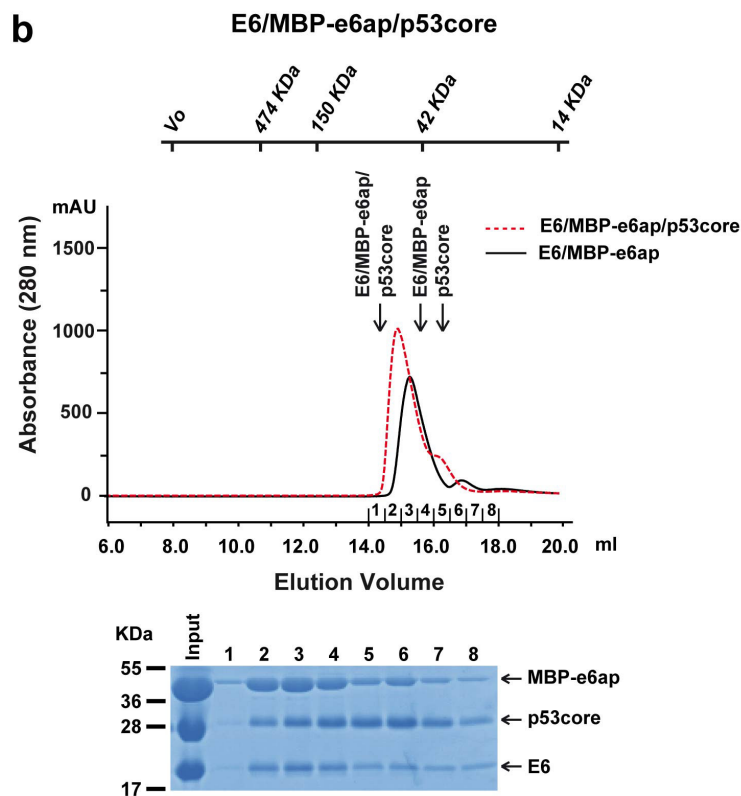
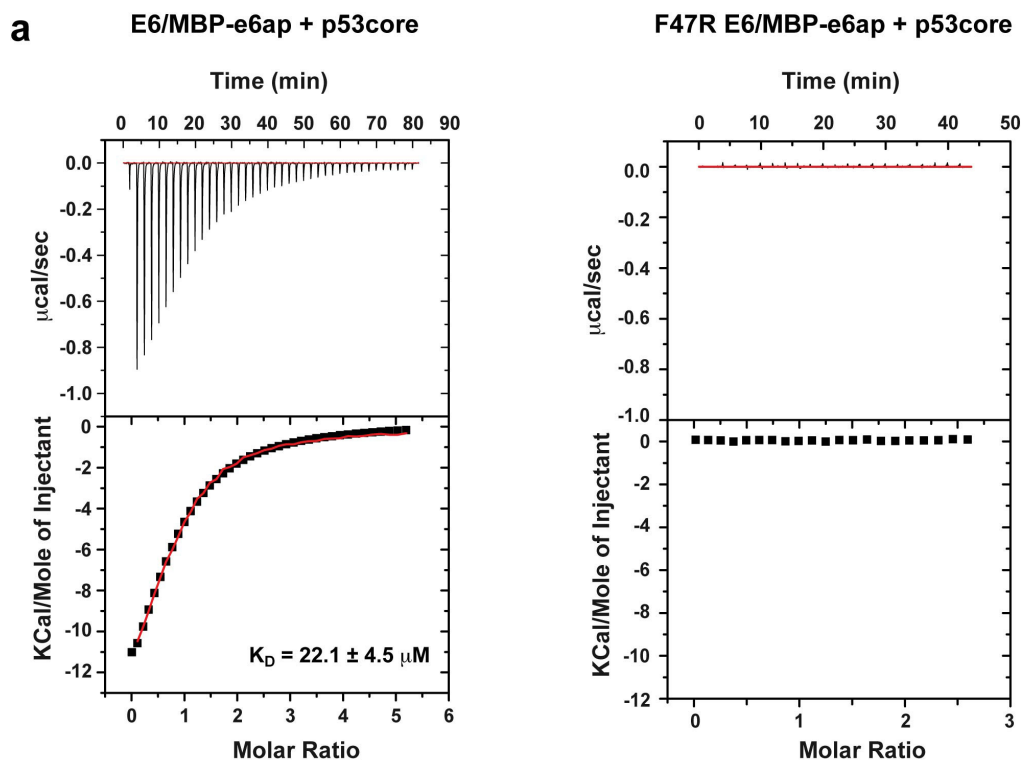
Extended Data Figure 1 | Constructs of the minimal HPV-16 E6/E6AP/p53 ternary complex. Green: E6AP residues 403–414 (e6ap 1–12) fused to the C terminus of MBP (pink) via a AAA linker (red); gold: the HPV-16

E6 4C/4S construct¹² (named E6 from hereon) comprising four cysteine to serine substitutions (cyan marks) that suppress aggregation mediated by disulfide cross-bridging; blue: p53 core domain (p53core).



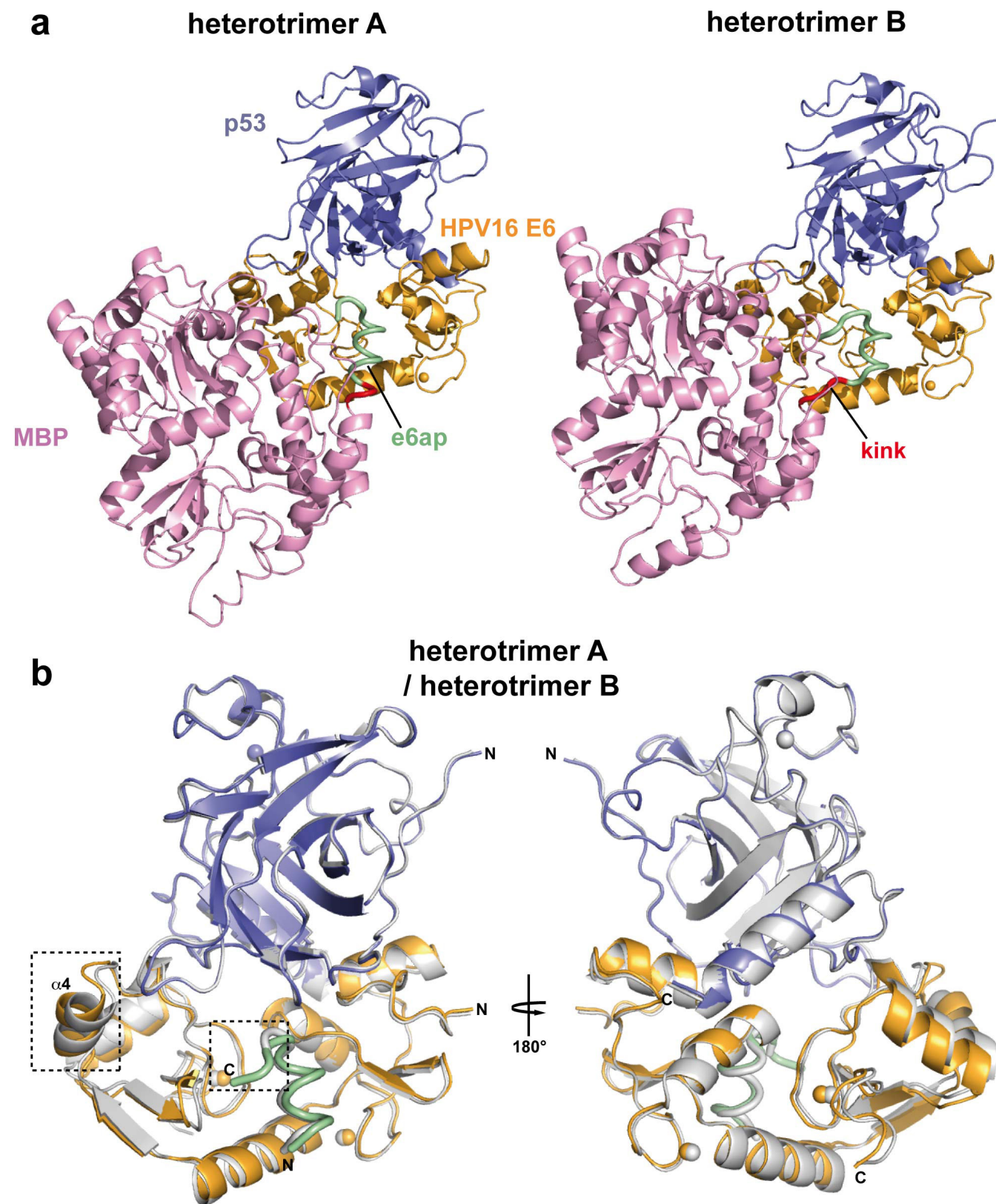
Extended Data Figure 2 | Assembly of the E6/MBP-e6ap heterodimer.
a, Dynamic light scattering (DLS) analysis of E6/MBP-e6ap samples. Histograms report the average hydrodynamic radii of the particles, whereas error bars indicate size polydispersity. Numbers above the histograms indicate molecular mass estimates assuming a spherical model. This analysis shows that binding to MBP-e6ap enhances the solubility of E6, which, in the unbound state, displays a solubility threshold of 50 μ M. However, in the case of E6/MBP-e6ap samples (grey histograms), particle size increases with raising concentration. By contrast, introduction of the F47R mutation in E6 (green histograms) stabilizes particle size to values close to what is expected for a simple heterodimer (\sim 60 kDa). Therefore

we conclude that, despite the increase in solubility, E6/MBP-e6ap still undergoes weak self-association via the E6N region hosting Phe47 (see ref. 12 for further discussion). **b**, Gel filtration analysis of E6/MBP-e6ap samples. The elution volumes for molecular size markers are reported on top of the figure. The expected elution volumes of a simple E6/MBP-e6ap heterodimer (1 \times , 60 kDa) and of a dimer of heterodimers (2 \times , 120 kDa) are indicated. Note the relative small shift in the elution volumes of the different samples compared with the differences in the hydrodynamic radii (**a**). This suggests that oligomers of the E6/MBP-e6ap heterodimer are rather weak and dissociate on the gel filtration column. See also Supplementary Methods.



Extended Data Figure 3 | Interaction of p53core with preformed E6/MBP-e6ap heterodimer. **a**, Isothermal titration calorimetry experiments were performed by titrating increasing amounts of p53core into E6/MBP-e6ap heterodimer samples adjusted to a concentration of $45 \mu\text{M}$, which limits heterodimer oligomerization. Note that the F47R mutation in E6 abolishes binding to p53core. See also Extended Data Table 1. **b**, Top: comparison of gel filtration elution profiles of E6/MBP-e6ap/p53core (red dashed line) versus E6/MBP-e6ap samples (black line). Both samples were

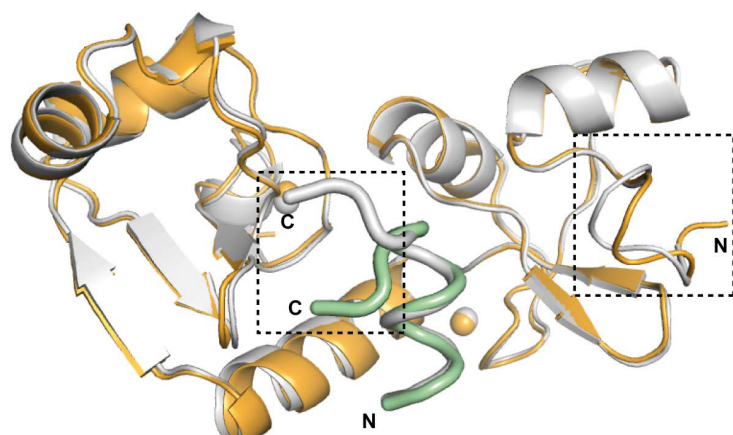
adjusted to a concentration of $250 \mu\text{M}$ before loading onto the gel filtration column. The expected elution volumes for p53core (24 kDa), monomeric E6/MBP-e6ap heterodimer (60 kDa) and E6/MBP-e6ap/p53core ternary complex (80 kDa) are indicated. Bottom: SDS-PAGE analysis of fractions comprising the elution peak of the ternary complex. Note the significant shift in the elution volumes of the main peak in the two chromatograms, indicating formation of a ternary complex. See also Supplementary Methods.



Extended Data Figure 4 | Comparison of the structures of the two E6/MBP-e6ap/p53core heterotrimers (heterotrimers A and B) observed in the asymmetric unit. Green: e6ap fused to the C terminus of MBP (pink) via the AAA linker (red); gold: HPV-16 E6; blue: p53core; spheres: zinc atoms. **a**, Different orientation of MBP in the two heterotrimers, which results from the different conformations of the two AAA linkers. **b**, Superposition of the E6/e6ap/p53 regions of heterotrimers A (green/

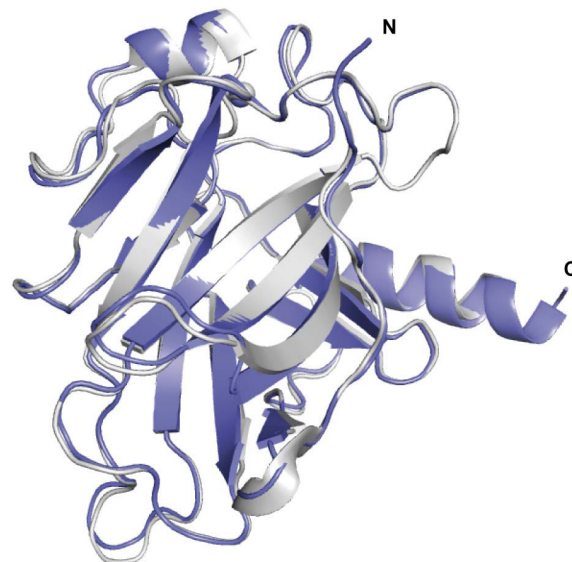
gold/blue) and B (grey). The backbone r.m.s.d. for the E6/e6ap/p53 regions of the two heterotrimers was calculated by aligning backbone atoms of residues (1) 12–136 of HPV-16 E6, (2) 371–379 of e6ap and (3) 109–191 of p53core and found to correspond to 0.9 Å. Regions displaying significant differences are boxed. These regions are the ill-defined C-terminal region of the e6ap peptide (see also Extended Data Fig. 6) and the $\alpha 4$ helix of E6, which is not involved in the E6–p53 interface.

HPV16 E6 / e6ap (heterotrimer A)
HPV16 E6 / e6ap (heterodimer, 4GIZ)

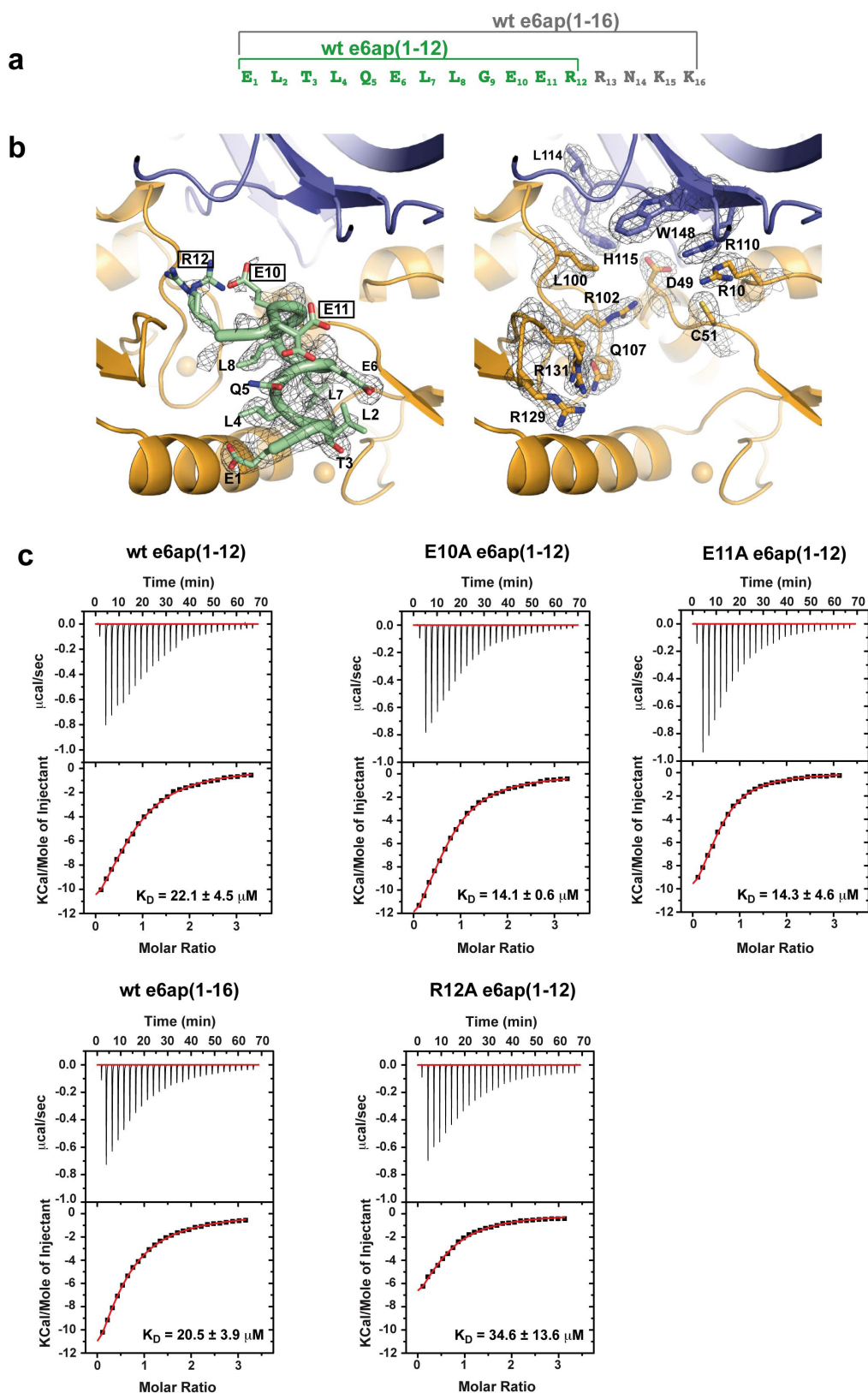


Extended Data Figure 5 | The structures of the E6/e6ap and p53core subunits of the ternary complex are superimposable with previously solved structures of the E6/e6ap heterodimer and p53core. Left: superposition of the previously solved E6/e6ap heterodimer (grey)⁸ onto the E6/e6ap subunit of the ternary complex (heterotrimer A, gold/green)

p53 core (heterotrimer A)
p53 core (unbound, 2OCJ)

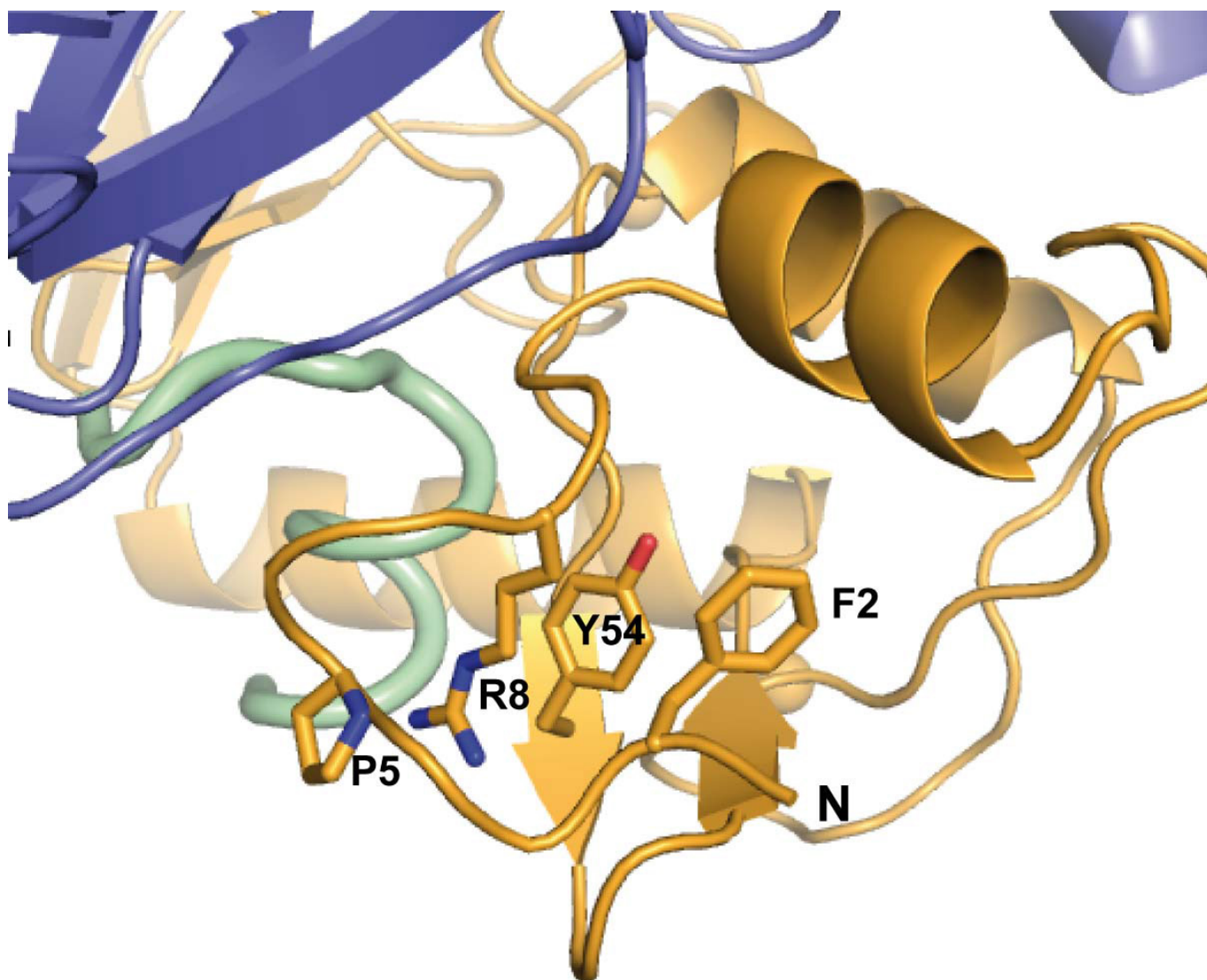


determined here. Dashed lines highlight regions of conformational change, namely the N terminus of HPV-16 E6 and the C terminus of the e6ap peptide. Right: superposition of previously solved p53core in the unbound state (heterotrimer A, grey)³⁷ onto the p53core subunit of the ternary complex bound (blue).



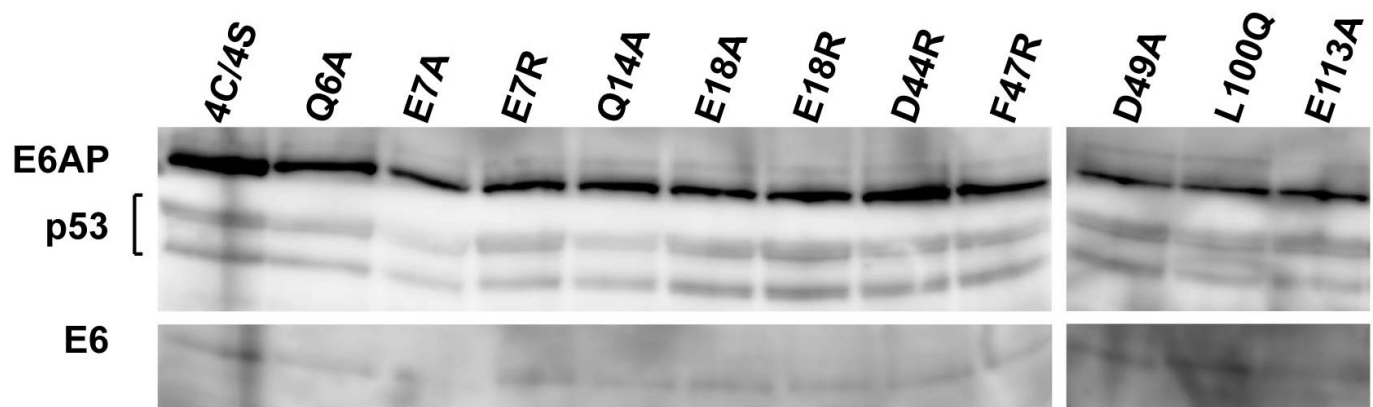
Extended Data Figure 6 | Contributions of the e6ap C-terminal region to ternary complex interface. **a**, Sequence of wild-type e6ap peptides used in the study. Green: wild-type e6ap(1–12) corresponding to the peptide used for crystallization of the ternary complex. Grey: wild-type e6ap(1–16) containing a four-amino-acid C-terminal extension. **b**, Electron density ($2F_o - F_c$) map (heterotrimer A) contoured at 1 s level for e6ap (green, left) and for selected E6 (orange) and p53 (blue) interface residues in the proximity of the C terminus of e6ap (right). Note the lack of electron density data for e6ap residues Glu10, Glu11 and Arg12. Two

conformations for the side chains Glu10 and Arg12 of e6ap and Arg131 of E6 are proposed in the model. **c**, Isothermal titration calorimetry curves showing the interactions of p53core with pre-formed E6/e6ap complexes bearing either point mutations within e6ap C terminus (E10A, E11A and R12A) or the wild-type e6ap(1–16) peptide construct. Note that neither the e6ap point mutations nor the e6ap C-terminal boundary affect the interaction between the E6/e6ap heterodimer and p53core. See also Extended Data Table 1.

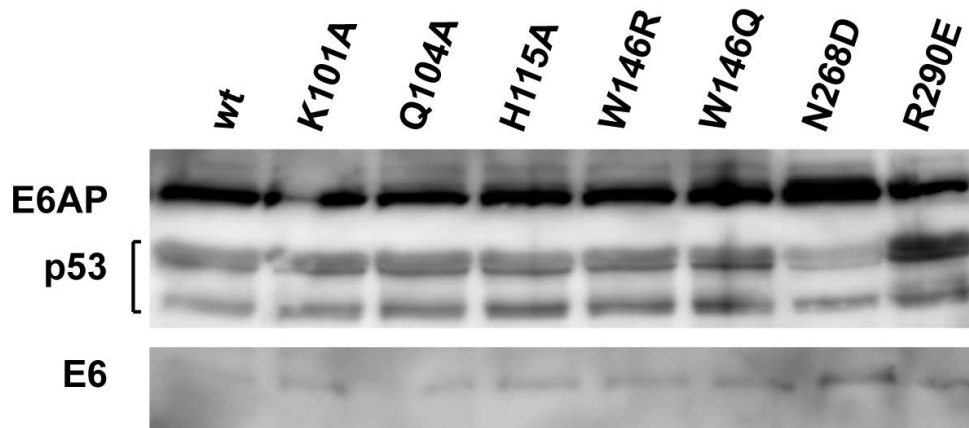


Extended Data Figure 7 | Interactions mediated by hrn-HPV-conserved residues shaping the conformation of the N-terminal arm of HPV-16 E6. Whereas Phe2 contributes to tethering of the N-terminal region to the core of the E6N domain, residues Pro5, Arg8 and Tyr54 are involved in a triple stacking interaction.

HPV16 E6 mutants



p53 mutants



Extended Data Figure 8 | Expression levels of HPV-16 E6, E6AP and p53 proteins in triple pull-down assays. HEK293T were transfected with 1 μ g of pSPICA-N1 plasmid (expressing E6AP fused to the C terminus of the *G. princeps* luciferase Gluc1 fragment), 1 μ g pSPICA-N2 plasmid (expressing the E6 fused to the C terminus of the *G. princeps* luciferase

Gluc2 fragment) and 1 μ g of BioEase plasmid (expressing N-terminally biotinylated p53). Cell lysates were resolved by SDS-PAGE electrophoresis and E6, E6AP and p53 fusion proteins detected by western blotting. For gel source data, see Supplementary Fig. 1.

Extended Data Table 1 | Thermodynamic parameters of p53core binding to preformed E6/e6ap complex

<i>e6ap</i>	[*] <i>N</i>	[†] <i>c</i> _{min} / <i>c</i> _{max}	<i>K_D</i> (μ M)	ΔH (kcal/mol)	<i>T</i> ΔS (kcal/mol)	ΔG (kcal/mol)
wt e6ap(1-12)	0.826±0.016	1.8/2.4	22.1±4.5	-17.4±1.5	-0.926±0.140	-16.5
wt e6ap(1-16)	0.516±0.046	1.8/2.7	20.5±3.9	-19.1±2.7	-1.100±0.299	-18.0
E10A e6ap(1-12)	0.827±0.080	3.1/3.3	14.1±0.6	-17.1±0.5	-0.878±0.051	-16.2
E11A e6ap(1-12)	0.679±0.146	2.6/4.0	14.3±4.6	-13.8±0.1	-0.599±0.012	-13.2
R12A e6ap(1-12)	0.719±0.111	1.0/2.2	34.6±13.6	-13.9±2.7	-0.651±0.250	-13.2

All isothermal titration calorimetry experiments were performed at 25 °C.

^{*}*N* refers to the molar ratio of p53core–E6/e6ap complex.

[†]The *c* value is defined as: $c = \eta[\text{E6/e6ap}]/K_D$.

Extended Data Table 2 | Data collection and refinement statistics

E6/e6ap/p53	
Data collection	
Space group	P 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.15 129.41 82.26
α , β , γ (°)	90, 92.4, 90
Resolution (Å)	50-2.25 (2.33-2.25) *
<i>R</i> _{sym} or <i>R</i> _{merge}	6.0 (68.4)
<i>I</i> / σ <i>I</i>	17.6 (1.8)
Completeness (%)	99.5 (99.6)
Redundancy	3.2 (3.1)
Refinement	
Resolution (Å)	50-2.25
No. reflections	77282
<i>R</i> _{work} / <i>R</i> _{free} (%)	19.4/24.6 (20.9/33.0)
No. atoms	12034
Protein	11593
Ligand/ion	84
Water	354
B-factors (Å ²) †	
Protein	50
MBP-e6ap (chains A, B)	50, 58
p53 (chains C, D)	43, 45
E6 (chains F, H)	49, 55
Water	47
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.37
Ramachandran	
No. residues ‡	1455
Favored (%)	95.94
Allowed (%)	3.99
Outliers (%)	0.07

*Highest resolution shell is shown in parenthesis.

†Values refer to occupancy-weighted average B-factors.

‡The total number of residues in the chain cannot be analysed: phi and psi angles cannot be analysed for terminal residues, non-standard residues or incompletely modelled main chain residues.

Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation

Bernhard C. Lechtenberg¹, Akhil Rajput², Ruslan Sanishvili³, Małgorzata K. Dobaczewska¹, Carl F. Ware², Peter D. Mace⁴ & Stefan J. Riedl¹

Ubiquitination is a central process affecting all facets of cellular signalling and function¹. A critical step in ubiquitination is the transfer of ubiquitin from an E2 ubiquitin-conjugating enzyme to a substrate or a growing ubiquitin chain, which is mediated by E3 ubiquitin ligases. RING-type E3 ligases typically facilitate the transfer of ubiquitin from the E2 directly to the substrate^{2,3}. The RING-between-RING (RBR) family of RING-type E3 ligases, however, breaks this paradigm by forming a covalent intermediate with ubiquitin similarly to HECT-type E3 ligases^{4–6}. The RBR family includes Parkin⁴ and HOIP, the central catalytic factor of the LUBAC (linear ubiquitin chain assembly complex)⁷. While structural insights into the RBR E3 ligases Parkin and HHARI in their overall auto-inhibited forms are available^{8–13}, no structures exist of intact fully active RBR E3 ligases or any of their complexes. Thus, the RBR mechanism of action has remained largely unknown. Here we present the first structure, to our knowledge, of the fully active human HOIP RBR in its transfer complex with an E2~ubiquitin conjugate, which elucidates the intricate nature of RBR E3 ligases. The active HOIP RBR adopts a conformation markedly different from that of auto-inhibited RBRs. HOIP RBR binds the E2~ubiquitin conjugate in an elongated fashion, with the E2 and E3 catalytic centres ideally aligned for ubiquitin transfer, which structurally both requires and enables a HECT-like mechanism. In addition, three distinct helix–IBR-fold motifs inherent to RBRs form ubiquitin-binding regions that engage the activated ubiquitin of the E2~ubiquitin conjugate and, surprisingly, an additional regulatory ubiquitin molecule. The features uncovered reveal critical states of the HOIP RBR E3 ligase cycle, and comparison with Parkin and HHARI suggests a general mechanism for RBR E3 ligases.

RBR E3 ligases are characterized by an extended RING domain (RING1) followed by an ‘in-between RING’ (IBR) domain and the catalytic domain, which is structurally an IBR domain but is commonly designated RING2 (Extended Data Fig. 1a, b)^{8–11,14}. HOIP, one of the most studied RBRs, is the key E3 ligase of the linear ubiquitin chain assembly complex (LUBAC). It is a prototypical RBR yet contains an extended RING2 domain that includes the LDD (linear ubiquitin chain determining domain, Extended Data Fig. 1c)^{5–7,14} and is thus denoted RING2L. The LDD enables the selective formation of linear ubiquitin linkages. The HOIP RBR is kept in an auto-inhibited state by the HOIP UBA domain, whose sequestration by the LUBAC constituent HOIL-1L activates HOIP to trigger, together with SHARPIN, NF- κ B signalling and other cellular processes^{5–7,15–20}. To obtain the first insight into an active RBR in a key catalytic complex, we generated a stable E2~ubiquitin conjugate (UbcH5B C85K~ubiquitin)²¹ and isolated its complex with HOIP RBR. The subsequent addition of free ubiquitin proved necessary for crystal formation, allowing us to solve the HOIP RBR/UbcH5B~ubiquitin transfer complex structure at 3.5 Å resolution (Fig. 1a; Extended Data Figs 2 and 3).

The asymmetric unit contains two HOIP RBR molecules interacting with two UbcH5B~ubiquitin conjugates and an additional ubiquitin or E2~ubiquitin conjugate, arranged in a swapped dimer configuration (Extended Data Fig. 3a). While this arrangement could have functional relevance, analysis of interfaces and biophysical examination (Extended Data Fig. 3b–f) indicate a monomeric assembly of the HOIP/E2~ubiquitin loading complex (Fig. 1), represented in the crystal structure by the RING1–IBR module (residues 699–852) from one HOIP molecule and the RING2L (residues 853–1,072) from the second HOIP molecule in the asymmetric unit. In this assembly, the RING1–IBR module forms an elongated arm-like unit (Fig. 2a) that

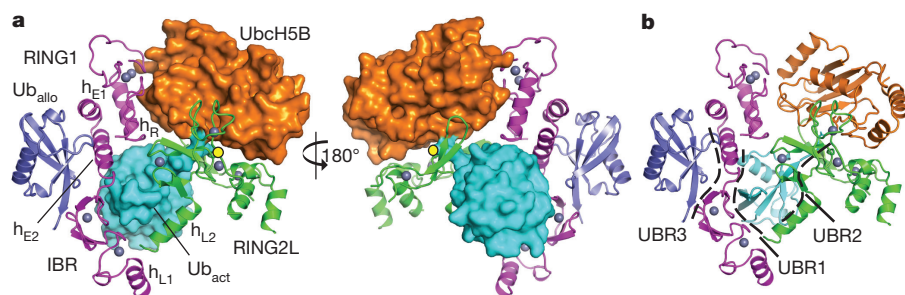


Figure 1 | Structure of the HOIP RBR/UbcH5B~ubiquitin transfer complex.

a, Structure with key elements annotated. The RBR RING1 and IBR together with two RING1 extension helices (h_{E1} , h_{E2}) form an arm-like unit (magenta; for domain annotations see Extended Data Fig. 1). The RING1–IBR and the catalytic RING2L (green) are connected by two linker helices (h_{L1} , h_{L2}) and together engage the E2~ubiquitin conjugate (UbcH5B~Ub_{act}; orange and cyan, respectively) positioning it for ubiquitin transfer onto the RBR catalytic cysteine (yellow circle). An allosteric ubiquitin molecule (Ub_{allo}, blue) binds to the RING1–IBR across from the activated ubiquitin. Zinc ions are shown as grey spheres; h_R denotes the RING1 helix. **b**, Location of the three ubiquitin-binding regions (UBRs, dashed lines) in the complex.

¹NCI-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. ²Infectious and Inflammatory Disease Center, Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. ³X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. ⁴Biochemistry Department, University of Otago, 710 Cumberland Street, Dunedin 9054, New Zealand.

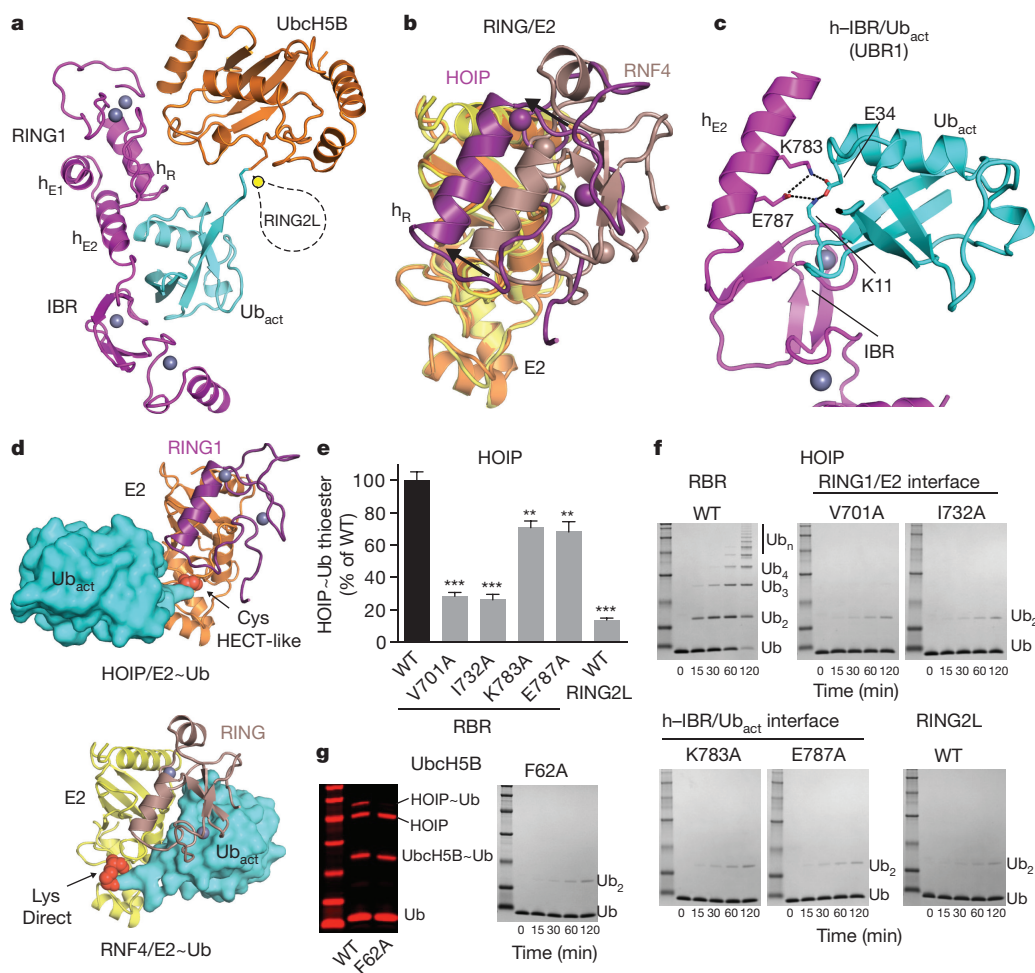


Figure 2 | The HOIP RING1-IBR coordinates the UbCH5B~ubiquitin conjugate in a bipartite manner tailored to a HECT-like mechanism.

a, Coordination of the UbCH5B~ubiquitin conjugate (orange and cyan) by the RING1-IBR (magenta). HOIP RING2L is indicated schematically. **b**, The HOIP RING1 coordinates the E2 in a shifted position compared to classic RINGs. Overlay of HOIP RING1/UbcH5B (magenta/yellow) and RNF4 RING/UbcH5A (brown/orange, PDB: 4AP4 (ref. 21)). **c**, RBR hE2 and IBR form UBR1 binding the activated ubiquitin. A central salt bridge system connects hE2 and ubiquitin. **d**, Comparison of E2~ubiquitin binding by HOIP RING1 (top) and a classic RING (bottom, RNF4 PDB).

4AP4 (ref. 21)) highlights the differences in E2~ubiquitin thioester positioning (red spheres). The directionality of the thioester attacking residue (active site cysteine of RBR in the HECT-like transfer and lysine of the substrate in RING-mediated transfer) is indicated. **e**, **f**, Quantitative thioester transfer (**e**) (mean activity \pm s.e.m. ($n = 3$); one-way ANOVA followed by Tukey's post-test; $^{**}P < 0.01$; $^{***}P < 0.001$; Supplementary Fig. 1) and linear polyubiquitination (**f**) assays of HOIP RBR wild type (WT) and interface mutants, and HOIP catalytic domain (RING2L). **g**, Thioester transfer and linear ubiquitin assays for UbCH5B wild type and RING1 interaction mutant F62A (Coomassie-stained bands in red).

together with the RING2L embraces the E2~ubiquitin conjugate in a clamp-like manner (Fig. 1a). This active HOIP RBR conformation is markedly different from previous structures of auto-inhibited RBRs (Extended Data Fig. 1d) and enables an astounding array of features inherent to the active RBR. Most notably, three distinct helix-IBR-fold motifs function as essential discrete ubiquitin-binding regions (UBR) (Fig. 1b).

The HOIP RING1/E2 interaction is tailored towards a HECT-like mechanism, setting it apart from classic RING E3 ligases. While RING/E2 interactions of both classic RING and RBR E3 ligases utilize similar surfaces (Extended Data Fig. 4a)^{21–26}, the position of the HOIP RING1 domain relative to the E2 is shifted compared to classic RING/E2 complexes (Fig. 2b). Therefore, the RBR RING1 and the E2 do not form a composite surface to bind the E2-conjugated activated ubiquitin (Ub_{act}, Extended Data Fig. 4b, c, e), which is key to the mechanism of classic RING E3 ligases^{21,24,27}. Instead, two extension helices (hE1, hE2) link the RBR RING1 to the IBR domain (Figs 1 and 2a)^{8–11}, and helix hE2 with the IBR forms an UBR (UBR1) that engages the activated ubiquitin (Fig. 2c and Extended Data Fig. 5a, b). UBR1 binds ubiquitin in a distinctive mode (mode 1) that utilizes a salt bridge system involving HOIP hE2 residues K783 and E787 and ubiquitin residues K11 and E34,

with further support from the HOIP IBR (Fig. 2c and Extended Data Fig. 5b). Thus, in HOIP the entire RING1-IBR arm mediates bipartite binding of the E2~ubiquitin conjugate, with RING1 binding the E2 and the hE2-IBR module binding activated ubiquitin (Fig. 2a). Sequence and structural comparisons with Parkin and HHARI suggest conservation of this mechanism among RBR E3 ligases (Extended Data Figs 4c and 5c). Importantly, the bipartite binding mode results in an elongated conformation of the E2~ubiquitin conjugate with its thioester linkage not suited for direct attack by the amine function of a substrate (Fig. 2d and Extended Data Fig. 4e–f). The consequence is an entirely different catalytic arrangement compared to classic RING-supported catalysis, as emphasized by the observed lack of effect of mutations in UbCH5B L104 and S108, two residues crucial for classic RING/E2 catalysis^{21,25,26} (Extended Data Fig. 4d). Instead, the E2~ubiquitin thioester is ideally positioned for transfer of the activated ubiquitin onto HOIP RING2L, thus both enabling and requiring a HECT-like mechanism. The importance of each interaction site is demonstrated in thioester transfer and polyubiquitination assays, where mutation of key RING1 and UbCH5B residues drastically impairs RBR activity (Fig. 2e–g). Single mutations in the hE2 salt bridge moderately diminish activity in thioester assays but dramatically impair polyubiquitination

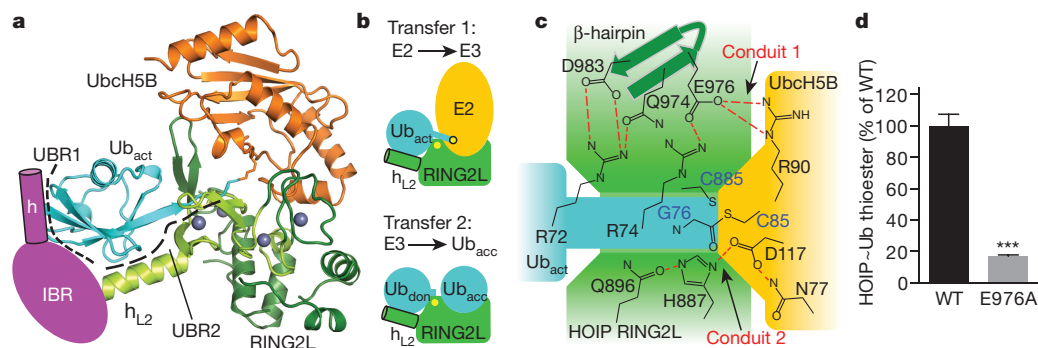


Figure 3 | Mechanism of E2~ubiquitin/HOIP RBR ubiquitin transfer. **a**, UbcH5B~ubiquitin conjugate bound to HOIP RING2L (RING2, light green and the linear ubiquitin chain determining domain (LDD) extension, dark green). UBR1 (schematic of helix-IBR, magenta) co-operates with UBR2 (comprising h_{L2} and IBR fold of RING2L) to bind the activated ubiquitin. UbcH5B interacts with RING2L in a region designated for the acceptor ubiquitin¹⁴ (displayed in **b**). **b**, Positioning of E2~Ub_{act} or

activity (Fig. 2e, f), indicating a cumulative effect due to a potential role of UBR1 in coordinating Ub_{act} in steps subsequent to its initial transfer to HOIP. However, removal of the salt bridge in HOIP/ubiquitin double mutants and mutation of the HOIP IBR/Ub_{act} interface cause the expected drastic reduction in thioester transfer activity (Extended Data Fig. 5d–f).

The other portion of the RBR/E2~ubiquitin embrace is centred around the catalytic HOIP RING2L (Figs 1a and 3a). Here a helix-IBR-fold motif consisting of helix h_{L2} from the IBR–RING2L linker and RING2L form a second UBR (UBR2) binding the activated ubiquitin (Fig. 3a and Extended Data Fig. 6a–c). UBR2, which is conserved in Parkin and HHARI (Extended Data Fig. 6d), uses a hydrophobic pattern in helix h_{L2} and RING2L to interact with the canonical (I44) and a second hydrophobic patch²⁸ of ubiquitin (Extended Data Fig. 6b–e). These interactions support the engagement of the ubiquitin R72/R74 di-Arg motif by polar residues, which is the key characteristic of UBR binding mode 2. This ultimately places the ubiquitin C terminus onto RING2 (Extended Data Fig. 6b, d) and thus the ubiquitin/E2-thioester linkage onto the RBR active site. A previous structure¹⁴ of the isolated HOIP RING2L with two ubiquitin molecules bound in a linear non-covalent arrangement mimics the final HOIP RING2L donor/acceptor ubiquitin transfer complex (Fig. 3b and Extended Data Fig. 6f). Remarkably, in this structure the donor ubiquitin adopts a position identical to the activated ubiquitin in HOIP UBR2 despite lacking the h_{L2} interaction (Extended Data Fig. 6f). This indicates that the binding mode of the activated ubiquitin observed in the HOIP/E2~ubiquitin complex persists from the E2~ubiquitin/E3 HOIP transfer complex to the HOIP~ubiquitin/acceptor ubiquitin transfer complex. The E2 portion of the E2~ubiquitin conjugate instead interacts with a region of RING2L that overlaps with that observed for the acceptor ubiquitin in the second transfer reaction (Fig. 3b and Extended Data Fig. 6f). Thus, the binding of E2~ubiquitin and of acceptor ubiquitin/substrate are mutually exclusive, requiring formation of a covalent HECT-like RBR~ubiquitin intermediate in the RBR E3 ligase cycle.

Importantly, the HOIP RBR/E2~ubiquitin complex structure lacks the spatial gap between the E2 and E3 catalytic centres that is frequently observed in HECT/E2 complex structures and that was also predicted for RBR/E2~ubiquitin transfer complexes^{8–11,28,29}. Thus, except for the ~3.5 Å spacer due to the C85 to lysine substitution in the E2~ubiquitin conjugate, the HOIP RBR/E2~ubiquitin structure accurately depicts the immediate transfer complex. Here the catalytic centres of HOIP RING2L and E2 come in close proximity via two contact conduits involving all three proteins (Fig. 3c and Extended Data Fig. 7a). The first conduit consists of ubiquitin R72, which interacts with D983 and Q974 in the β 5/6-hairpin of HOIP RING2L. Additionally, E976 in this hairpin mediates a salt bridge with ubiquitin

donor/acceptor ubiquitin (Ub_{don} or Ub_{acc}, respectively) onto RING2L. **c**, Ternary HOIP RBR/E2~ubiquitin catalytic transfer complex. Two main contact conduits (red dashes) position the RBR catalytic cysteine (C885) near the E2 C85-ubiquitin thioester linkage. **d**, HOIP~ubiquitin thioester formation assay with wild-type or E976A RBR, supports the HOIP/UbH5B/Ub link in conduit 1 (mean activity \pm s.e.m. ($n=3$); two-tailed unpaired Student's *t*-test; *** $P < 0.001$; Supplementary Fig. 1).

R74 and UbcH5B R90, thus facilitating interactions among all three proteins. The second conduit consists of catalytic residues of UbcH5B (N77, D117) and HOIP (H887, Q896)^{5,6,10,14,21–24,26}. These residues appear permissive to close proximity between the reaction centres, yet not crucial for transesterification because, for example, a H887A mutation does not affect the thioester transfer reaction^{10,14} (Extended Data Fig. 7b). Surprisingly, mutation to alanine of UbcH5B D117, a critical residue for classic RING-supported catalysis^{21,24}, enhances transesterification (Extended Data Fig. 7b), further underlining the vastly different catalytic mechanism of RBR E3 ligases. This finding also points to a trade-off in the E2 active site to support both classic

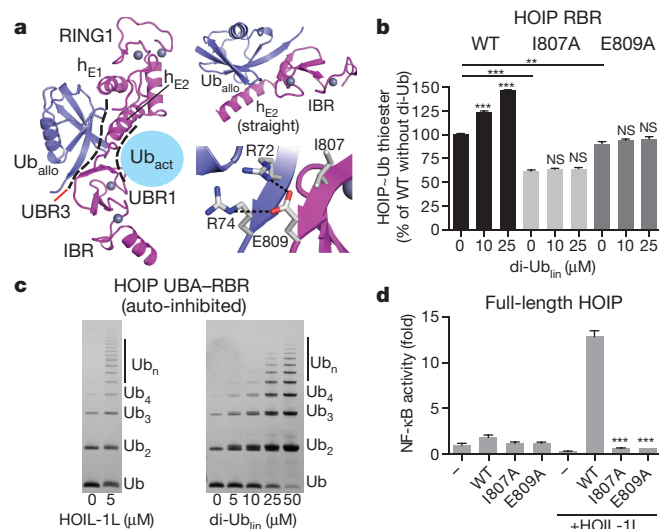


Figure 4 | An allosteric ubiquitin interacts with UBR3 in the RING1–IBR arm and is crucial for HOIP activity. **a**, An allosteric ubiquitin (Ub_{allo}, blue) binds to UBR3 across the activated ubiquitin. Left, overview depicting UBR3/Ub_{allo}. Top right, UBR3 h_{E2} –IBR/Ub_{allo} interaction. Bottom right, magnified view of the HOIP ubiquitin di-Arg binding motif (E809) anchoring a parallel ubiquitin/IBR β -sheet. **b**, Linear di-ubiquitin increases HOIP RBR activity. Thioester transfer assays of wild-type HOIP and UBR3 mutants (mean activity \pm s.e.m. ($n=3$); one-way ANOVA followed by Tukey's post hoc test; *** $P < 0.01$; *** $P < 0.001$; NS, not significant; Supplementary Fig. 1). **c**, Polyubiquitination assays showing release of HOIP UBA–RBR auto-inhibition by HOIL-1L or linear di-ubiquitin. **d**, Effect of wild-type HOIP or UBR3 mutants in NF- κ B reporter assays using HEK293T cells expressing full-length HOIP with or without HOIL-1L (mean activity \pm s.e.m. of three biological replicates each with three technical replicates; one-way ANOVA followed by Tukey's post hoc test; *** $P < 0.001$; Extended Data Fig. 9f).

RING- and HECT-type RBR E3 ligases. Notably, UbcH7, which is specialized for HECT-like E3 catalysis⁴, features a histidine instead of D117 (Extended Data Fig. 7a). Mutational analysis demonstrates a crucial role for conduit 1 and also indicates that the close proximity between the ubiquitin thioester (and thus C85 of UbcH5B) and HOIP catalytic cysteine C885 is the driving factor for E2/RBR E3 ubiquitin transfer (Fig. 3c, d and Extended Data Fig. 7). Analysis of Parkin and HHARI shows conservation of the conduits (Extended Data Fig. 7a). However, while Parkin and HHARI lack the β 5/6-hairpin indigenous to HOIP RING2L, they instead possess a pair of conserved polar residues in the RING2 active site loop that are capable of binding the di-Arg motif in conduit 1.

Surprisingly, our structure reveals that an additional allosteric ubiquitin molecule (Ub_{allo}) interacts with a third HOIP UBR. UBR3 is located in the RING1–IBR arm immediately across UBR1 and Ub_{act} (Figs 1b, 4a and Extended Data Fig. 8a, b). Ub_{allo} uses a binding mode similar to that of Ub_{act} with UBR2 (mode 2), characterized by hydrophobic interactions and a di-arginine binding clamp (Extended Data Fig. 8a). Ub_{allo} interacts with helix h_{E2} of the extended RING1 and with the IBR, and makes additional interactions with helix h_{E1} (Fig. 4a and Extended Data Fig. 8c). Through this binding, Ub_{allo} induces a 'straight' conformation of helix h_{E2} locking RING1 and IBR in their relative position, forming UBR1 to accommodate the activated ubiquitin (Fig. 4a and Extended Data Fig. 8b–d). Notably, UBR3 in the HOIP RBR/UbcH5B~ubiquitin complex binds linear di-ubiquitin ($K_d = 7 \mu M$) better than mono-ubiquitin ($K_d > 50 \mu M$) (Extended Data Fig. 9a). Pre-incubation of HOIP RBR with linear di-ubiquitin leads to improved binding of UbcH5B~ubiquitin (Extended Data Fig. 9b), emphasizing the allosteric function of UBR3. Accordingly, HOIP UBR3 I807A and E809A mutants show moderately decreased activity in thioester transfer assays but more pronounced effects in polyubiquitination assays, where linear di-ubiquitin/polyubiquitin are intrinsically produced (Fig. 4b and Extended Data Fig. 9c). Importantly, N- and C-terminally capped linear di-ubiquitin increases HOIP RBR thioester transfer activity in a dose-dependent manner, but cannot activate HOIP UBR3 mutants (Fig. 4b and Extended Data Fig. 9d). Moreover, the linear di-ubiquitin I44A mutant also fails to activate HOIP RBR (Extended Data Fig. 9d).

Excitingly, the interaction of Ub_{allo} with UBR3 is structurally similar to that recently reported for phospho-ubiquitin in a tethered complex with ΔUBL *Pediculus humanus* Parkin¹³ (Extended Data Fig. 8c–e). Binding of phospho-ubiquitin leads to a straight conformation of Parkin helix h_{E2} and an accompanying reorientation of RING1 and IBR, indicating a general role of UBR3 and ubiquitin in allosteric regulation of RBR proteins. Functionally, binding of phospho-ubiquitin activates Parkin by counteracting the auto-inhibitory function of the Parkin UBL domain (Extended Data Fig. 8c–e)¹³. In HOIP, the UBA domain exerts intramolecular auto-inhibition^{5,6}. While the structure of auto-inhibited HOIP is unknown, the structure of auto-inhibited HHARI shows binding of its UBA domain to a region analogous to UBR3 (Extended Data Fig. 8f–h)¹¹. To determine if linear di-ubiquitin can overcome HOIP auto-inhibition, we examined its effect on HOIP UBA–RBR. As expected, HOIP UBA–RBR alone exhibits low E3 activity but is activated by HOIL-1L (Fig. 4c). Notably, linear di-ubiquitin can also remove HOIP UBA auto-inhibition and at high concentrations allows the processive formation of polyubiquitin chains by HOIP UBA–RBR (Fig. 4c). Importantly, in HEK293T cells expressing full-length HOIP, the HOIP UBR3 I807A and E809A mutants fail to activate NF- κ B regardless of HOIL-1L expression, demonstrating an essential physiological role of UBR3 (Fig. 4d). Thus, UBR3 probably serves as a critical sensor of ubiquitin chains that regulates LUBAC function. Whether this role is tailored to linear ubiquitin chains or ubiquitin chains in general (Extended Data Fig. 9b, e) needs further investigation in the context of other LUBAC constituents and binding partners.

The features revealed by the HOIP RBR/E2~ Ub_{act} / Ub_{allo} complex structure provide the missing links in our understanding of these

enigmatic multidomain E3 ligases^{8–11,14} and yield a mechanistic model for the RBR E3 ubiquitin ligase cycle, as summarized in Extended Data Fig. 10. Furthermore, the conservation of key mechanistic features in HOIP, HHARI, Parkin and other RBRs (Supplementary Data 2) underlines the general nature of the catalytic RBR cycle revealed in this study.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 July; accepted 3 December 2015.

Published online 20 January 2016.

- Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).
- Metzger, M. B., Pruneda, J. N., Klevit, R. E. & Weissman, A. M. RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim. Biophys. Acta* **1843**, 47–60 (2014).
- Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nature Struct. Mol. Biol.* **21**, 301–307 (2014).
- Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
- Smit, J. J. et al. The E3 ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension. *EMBO J.* **31**, 3833–3844 (2012).
- Stieglitz, B., Morris-Davies, A. C., Koliopoulos, M. G., Christodoulou, E. & Rittinger, K. LUBAC synthesizes linear ubiquitin chains via a thioester intermediate. *EMBO Rep.* **13**, 840–846 (2012).
- Kirisako, T. et al. A ubiquitin ligase complex assembles linear polyubiquitin chains. *EMBO J.* **25**, 4877–4887 (2006).
- Riley, B. E. et al. Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nature Commun.* **4**, 1982 (2013).
- Trempe, J. F. et al. Structure of parkin reveals mechanisms for ubiquitin ligase activation. *Science* **340**, 1451–1455 (2013).
- Wauer, T. & Komander, D. Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J.* **32**, 2099–2112 (2013).
- Duda, D. M. et al. Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. *Structure* **21**, 1030–1041 (2013).
- Kumar, A. et al. Disruption of the autoinhibited state primes the E3 ligase parkin for activation and catalysis. *EMBO J.* **4**, 2506–2521 (2015).
- Wauer, T., Simicek, M., Schubert, A. & Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* **524**, 370–374 (2015).
- Stieglitz, B. et al. Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature* **503**, 422–426 (2013).
- Gerlach, B. et al. Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* **471**, 591–596 (2011).
- Ikedo, F. et al. SHARPIN forms a linear ubiquitin ligase complex regulating NF- κ B activity and apoptosis. *Nature* **471**, 637–641 (2011).
- Tokunaga, F. et al. SHARPIN is a component of the NF- κ B-activating linear ubiquitin chain assembly complex. *Nature* **471**, 633–636 (2011).
- Yagi, H. et al. A non-canonical UBA-UBL interaction forms the linear-ubiquitin-chain assembly complex. *EMBO Rep.* **13**, 462–468 (2012).
- Sasaki, K. & Iwai, K. Roles of linear ubiquitinylation, a crucial regulator of NF- κ B and cell death, in the immune system. *Immunol. Rev.* **266**, 175–189 (2015).
- Tokunaga, F. et al. Involvement of linear polyubiquitylation of NEMO in NF- κ B activation. *Nature Cell Biol.* **11**, 123–132 (2009).
- Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120 (2012).
- Mace, P. D. et al. Structures of the cIAP2 RING domain reveal conformational changes associated with ubiquitin-conjugating enzyme (E2) recruitment. *J. Biol. Chem.* **283**, 31633–31640 (2008).
- Plechanovová, A. et al. Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nature Struct. Mol. Biol.* **18**, 1052–1059 (2011).
- Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nature Struct. Mol. Biol.* **19**, 876–883 (2012).
- Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
- Pruneda, J. N. et al. Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942 (2012).
- Branigan, E., Plechanovová, A., Jaffray, E. G., Naismith, J. H. & Hay, R. T. Structural basis for the RING-catalyzed synthesis of K63-linked ubiquitin chains. *Nature Struct. Mol. Biol.* **22**, 597–602 (2015).
- Kamadurai, H. B. et al. Insights into ubiquitin transfer cascades from a structure of a UbcH5B~ubiquitin-HECT(NEDD4L) complex. *Mol. Cell* **36**, 1095–1102 (2009).
- Huang, L. et al. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2–E3 enzyme cascade. *Science* **286**, 1321–1326 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank A. Bobkov (SBP Protein Analysis Facility) for performing ITC and AUC experiments, M. Petroski (SBP) for providing UbH5B and Cdc34 constructs, J. Badger (DeltaG technologies) for assistance in model evaluation and E. Pasquale (SBP) for help with manuscript writing. This work was supported by NIH grant R01AA017238 and institutional funding (S.J.R.), an EMBO Long-term Postdoctoral Fellowship (B.C.L.), a Rutherford Discovery Fellowship from the New Zealand government administered by the Royal Society of New Zealand (P.D.M.) and NCI Cancer Center Support Grant P30CA030199 (SBP Protein Analysis Core Facility). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006).

Author Contributions B.C.L. designed and carried out all experiments (except for the cell-based experiments), including crystallization, structure solution and refinement, and wrote the manuscript. M.K.D. expressed proteins and performed initial purification. P.D.M. participated in early stages of the study, structure solution and writing of the manuscript. R.S. collected and processed diffraction data. A.R. performed the HEK293T cell experiments under the supervision of C.F.W. S.J.R. oversaw and actively participated in all steps of the study and wrote the manuscript.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 5EDV. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.R. (sriedl@SBPdiscovery.org).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Constructs. Human HOIP and HOIL-1L cDNA were purchased from Open Biosystems (cloneIDs 4653017 and 3877587, respectively). HOIP RBR (residues 696–1,072), HOIP RING2L (residues 853–1,072) and full-length HOIL-1L were cloned into the pET-NKI-6xHis-3C-LIC vector³⁰ coding for an N-terminal 6×His tag with a 3C protease cleavage site. HOIP UBA–RBR (residues 475–1,072) was cloned into a pET-NKI-6×His-eGFP-3C-LIC vector that codes for a 3C-cleavable His-tagged enhanced green fluorescent protein (eGFP) followed by the HOIP sequence. Human UbcH5B and Cdc34 DNA were a gift from M. Petroski. Coding sequences for UbcH13 and Uev1a were extracted out of a human cDNA library (Agilent Megaman). For crystallization, UbcH5B (residues 2–147) with the mutations S22R (to prevent backside ubiquitin binding³¹) and C85K (to enable covalent ubiquitin linkage²¹) was cloned into the pET-NKI-6×His-3C-LIC vector. UbcH5B without S22R and C85K mutations (used for enzymatic assays), Cdc34, UbcH13 and Uev1a were cloned into the same vector. Untagged mono-ubiquitin with native N and C termini, used for crystallization and linear ubiquitination assays, was cloned into the pET29 vector (Novagen) using NdeI/XhoI restriction sites. N-terminally blocked mono-ubiquitin used for thioester assays was cloned in the pET-NKI-6×His-3C-LIC vector. Untagged linear di-ubiquitin was cloned with overlap extension PCR and ligated into the pET29 vector (Novagen) using NdeI/XhoI restriction sites. N- and C-terminally blocked di-ubiquitin with a N-terminal His tag and a C-terminal Ala–Ser sequence was cloned into the pET-NKI-6×His-3C-LIC vector. Human ubiquitin-activating enzyme E1 (Ube1) was cloned into a pET28 vector resulting in an N-terminal His tag. For NF-κB assays full-length HOIP with an N-terminal Flag tag and HOIL-1L with an N-terminal myc tag were cloned into pcDNA3.1(+) (Invitrogen) using EcoRI/NotI restriction sites. Mutations in UbcH5B, ubiquitin and HOIP were introduced using standard site-directed mutagenesis techniques.

Protein expression and purification. All proteins were expressed in BL21(DE3) *E. coli* after induction with 0.5 mM IPTG overnight at 20°C. For expression of HOIP and HOIL-1L constructs, 0.5 mM ZnCl₂ was added to the cultures before induction. Bacteria were harvested by centrifugation, lysed by addition of lysozyme and sonication in the presence of protease inhibitors (PMSF and leupeptin) and DNase. Lysates were cleared by centrifugation and His-tagged proteins were initially purified using Ni-NTA agarose (Qiagen). For HOIP RBR used for crystallization, and UbcH5B, Cdc34, UbcH13, Uev1a, wild-type ubiquitin to generate K48-linked di-ubiquitin and HOIL-1L His tags were removed by addition of 3C protease overnight at 4°C. HOIP RBR and HOIL-1L were further purified using Superdex 200 10/300 GL or HiLoad 16/600 Superdex 200 pg size-exclusion chromatography columns (GE Healthcare) equilibrated in protein buffer (10 mM HEPES pH 7.9, 100 mM NaCl). UbcH5B used for biochemical assays was further purified on a Superdex 75 10/300 GL size-exclusion chromatography column (GE Healthcare) equilibrated in protein buffer. HOIP mutants for activity assays, and Cdc34, UbcH13 and Uev1a were desalted into protein buffer directly after Ni-NTA purification using PD MidiTrap G-25 desalting columns (GE Healthcare). Ube1 for biochemical assays was further purified using ion-exchange chromatography (Source Q) in 10 mM HEPES pH 7.9, 10 mM NaCl and eluted with a gradient from 10–500 mM NaCl. N-terminally His-tagged (di-)ubiquitin was purified using Ni-NTA as described above followed by size-exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare) equilibrated in protein buffer or buffer exchange into protein buffer using PD MidiTrap G-25 desalting columns. To purify untagged mono- or di-ubiquitin, 0.5 mM EDTA and 100 mM sodium acetate pH 4.5 were added to the bacterial lysates and lysates were cleared by centrifugation, diluted sevenfold with 50 mM sodium acetate pH 4.5 and applied to a Source S 10/100 ion exchange column (GE Healthcare) equilibrated in 50 mM sodium acetate pH 4.5. Ubiquitin was eluted with a 0–500 mM NaCl gradient and further purified by size-exclusion chromatography on a Superdex 75 10/300 GL column (GE Healthcare) equilibrated in protein buffer. His-eGFP-HOIP was purified using size-exclusion chromatography as described for HOIP RBR, followed by 3C cleavage and removal of His-eGFP via a second round of size-exclusion chromatography. All proteins were generally flash frozen in liquid nitrogen in small aliquots and stored at –80°C.

UbcH5B~ubiquitin linkage. UbcH5B~ubiquitin linkage was performed based on published methods²¹. Briefly, Ube1, UbcH5B(S22R/C85K) and ubiquitin were mixed and buffer exchanged into 50 mM Tris pH 10, 150 mM NaCl using PD-10 desalting columns (GE Healthcare). 10 mM MgCl₂, 5 mM ATP and 1 mM TCEP were added and the protein solution was incubated at 37°C for 16 h. The completeness of the reaction was monitored using SDS–PAGE and covalently linked UbcH5B~ubiquitin was purified from unreacted proteins and Ube1 using a

Superdex 75 10/300 GL size-exclusion chromatography column (GE Healthcare) equilibrated in protein buffer.

HOIP RBR/UbcH5B~ubiquitin complex formation. HOIP RBR was mixed with a 1.3-fold molar excess of UbcH5B~ubiquitin and applied to a Superdex 200 10/300 GL size-exclusion chromatography column equilibrated in protein buffer. Complex formation and purity was confirmed using SDS–PAGE, and complex containing fractions were pooled and concentrated to ~12 mg ml^{–1} for crystallization.

HOIP/UbcH5B~ubiquitin/ubiquitin crystallization. Crystallization was performed using the vapour diffusion technique in sitting drop MRC 96-well plates (Molecular Dimensions). Initial crystals were obtained mixing HOIP/UbcH5B~ubiquitin complex solution with an equimolar amount of free ubiquitin in the Morpheus Screen (Molecular Dimensions). Subsequently, 2 µl of the protein complex were mixed with 0.6 µl reservoir solution (0.1 M Morpheus Buffer 3 pH 8.5 (Tris/Bicine), 0.12 M Morpheus Alcohols Mix (0.02 M each of 1,6-hexanediol; 1-butanol; 1,2-propanediol (racemic); 2-propanol; 1,4-butanediol; 1,3-propanediol), 30% Morpheus P550MME_P20K mix (20% PEG550MME, 10% PEG20K) and 8% glycerol) in MRC 48-well plates (Molecular Dimensions). Crystals appeared after about one week at 12°C and were cryo-cooled, and evaluated on a rotating anode X-ray generator (Rigaku FR-E superbright). Seeding and dehydration of the crystals was performed to improve crystal diffraction. For successful dehydration, reservoir was slowly added to the protein drop (3 × 0.5 µl within ~2 h) and subsequently equilibrated overnight at 12°C against a reservoir solution with increased P550MME_P20K concentration by adding 11 µl 60% Morpheus P550MME_P20K stock solution to 50 µl reservoir solution. The new reservoir solution was then slowly added to the protein drop (3 × 0.5 µl, followed by 2 × 1 µl with removal of 1 µl each in the last steps). After further overnight equilibration, crystals were harvested from the drop and directly cryo-cooled in a cryogenic nitrogen stream at 100 K. Crystals diffracted in-house to 4–6 Å. Complete diffraction data were measured at 100 K at beamline 23ID-D of the General Medical Sciences and Cancer Institutes Structural Biology Facility at the Advanced Photon Source (GM/CA @ APS), Argonne National Laboratory. Despite their size (common dimensions of ~200 × 140 × 100 µm³) crystals exhibited substantial inhomogeneity resulting in split and smeared diffraction spots. Using raster scans³², a suitable region for data collection could be identified at the edge of the crystal. Using a small (20 µm diameter) beam, split spots could be separated to allow reliable indexing and integration. Utilization of a small beam necessitated higher flux to retain reliable diffraction. To mitigate the radiation damage, the total dose was distributed over a 100-µm stretch of the crystal by using the ‘helical’ mode of ‘vector’ data collection as implemented in JBluCE³³. Data were measured at 1.282 Å wavelength with a Pilatus3 6M pixel array detector with a 1-mm-thick sensor (Dectris).

Data processing and structure solution. Data were collected from a single crystal and indexed, integrated and scaled in XDS/XSCALE³⁴. Data were further processed using AIMLESS³⁵ from the CCP4 suite³⁶ with a resolution cut-off of 3.48 Å, resulting in an $\langle \frac{I}{\sigma} \rangle = 1.6$ and $CC1/2 = 0.648$ in the highest resolution shell. Phasing was carried out in Phaser³⁷ using an MR-SAD protocol as implemented in PHENIX³⁸. For this, independent molecular replacement searches were initially performed for the RING2L domain of HOIP (from PDB: 4LJP (ref. 14)), UbcH5B (from PDB: 3A33 (ref. 39)), and ubiquitin (from PDB: 4LJP (ref. 14)) with the four C-terminal residues deleted. Various ambiguous solutions were identified that could not be separated, and Zn²⁺ sites could not be identified using MR-SAD due to incompleteness of resultant models. However, manual inspection revealed that some MR solutions contained ubiquitin oriented near identically to the symmetry-related donor ubiquitin observed in the HOIP RING2L/ubiquitin-ubiquitin transfer complex (PDB: 4LJP (ref. 14)). Based on this observation, a trimmed search model was created that contained a complex of the core of HOIP RING2L (with residues 906–924 and 949–999 removed) and C-terminally truncated ubiquitin. An MR search using this model found a single solution for two copies of the complex. After successful iterative searches for two UbcH5B molecules and two further ubiquitin molecules, MR-SAD using Phaser identified 15 distinct Zn²⁺ sites including the known Zn²⁺ sites in the RING2L domain of HOIP. Further molecular replacement in Phaser using a single unit of the initial HOIP RING2L/UbcH5B~ubiquitin complex (without the additional second ubiquitin), and the NMR structure of HOIP IBR (zinc atoms removed, deposited in Protein Data Bank⁴⁰ under PDB accession number 2CT7, unpublished) correctly placed a single HOIP IBR domain, which was then manually copied to the other NCS-related HOIP in the asymmetric unit. For molecular replacement of the HOIP RING1, Sculptor⁴¹ was used to generate a search model based on the structure of the RING1 domain of HHARI (PDB: 4KBL (ref. 11)). However, Phaser was not able to correctly place this domain, probably owing to the low sequence conservation of only 27% identity. However, since mutational analysis of HOIP suggested that the RING/E2 interaction is preserved between RING-type E3 ligases and RBR-type E3 ligases⁵, we overlaid the

E2 of the published RNF4-RING/UbcH5A~ubiquitin structure (PDB: 4AP4 (ref. 21)) with the E2 in our structure and then used this overlay to add the RING1 model generated by Sculptor. This overlay placed the HOIP RING1 Zn²⁺-coordinating residues near the last remaining free Zn²⁺ ions found earlier by Phaser MR-SAD, indicating correct placement of the RING1 domain. In the final round of molecular replacement, the two additional ubiquitin (Ub_{allo}) molecules were reinstated at the RING1-IBR interface. At this stage, Refmac⁴² was used for refinement using settings optimized for low-resolution refinement⁴³ including 'jelly body refinement' and Babinet scaling. ProSMART⁴⁴ was used to generate external restraints against high-resolution structures (PDB: 4LJO (ref. 14) for HOIP RING2L and ubiquitin, and PDB: 2ESK (ref. 45) for UbcH5B). After this, clear extra electron density became visible for the unmodelled helical linker regions of the RING1-IBR and IBR-RING2L transitions and for other regions omitted in the initial search models. Further model building and refinement was manually performed in Coot⁴⁶ and Refmac. During refinement additional clear positive difference map electron density became visible and Phaser was used to place one additional UbcH5B molecule (UbcH5B_{add}) into this density. TLS restraints were generated using the TLSMD server⁴⁷ and NCS restraints were used throughout refinement. One overall B-factor was refined in Refmac. In later rounds of refinement the PDB_REDO server⁴⁸ was used for refinement optimization and MolProbity⁴⁹ was used for structure validation. Data processing and refinement statistics are summarized in Extended Data Fig. 2b. Ramachandran statistics were calculated using MolProbity and 94.8% of all residues are in favoured regions, 4.9% in allowed regions and 0.3% are outliers. The final structure has a MolProbity score of 1.75 (100th percentile). In the final structure the two HOIP RBR molecules (see also Extended Data Fig. 3) are defined by electron density from residues 699 to 707, 711 to 948, 969 to 991, and 996 to 1,011 (chain A) and 699 to 754, 760 to 957, 967 to 1,015, 1,019 to 1,035 and 1,054 to 1,066 (chain B). The catalytic UbcH5B~ubiquitin conjugates are defined from UbcH5B residues 3 to 147 and ubiquitin residues 1 to 76 (chains C and E), and UbcH5B residues 2 to 147 and ubiquitin residues 1 to 76 (chains D and F). The allosteric ubiquitin chains (chains G and H) are defined from residues 1 to 76 and the additional UbcH5B (chain I) is defined from residues 2 to 146. PHENIX was used to calculate simulated annealing (SA) composite omit maps and feature enhanced maps (FEM). All molecular figures were prepared in PyMOL (Schrödinger, LLC).

K48-linked and K63-linked ubiquitin chain formation. K48-linked and K63-linked ubiquitin chains were formed through a linkage-specific enzymatic reaction using Cdc34 and UbcH13/Uev1a E2 ubiquitin-conjugating enzymes, respectively, as described in the literature⁵⁰. Ubiquitin chains were separated using ion-exchange chromatography as described above for purification of mono-ubiquitin. Purified K48-linked di-ubiquitin was directly desalted into protein buffer using PD-10 desalting columns, whereas K63-linked di-ubiquitin was further purified on a Superdex 75 10/300 GL size-exclusion chromatography column equilibrated in protein buffer. Native ubiquitin without additional residues was used to generate di-ubiquitin chains for ITC experiments, whereas N-terminally blocked ubiquitin was used to form K48-linked di-ubiquitin for testing allosteric activation of HOIP RBR.

Linear polyubiquitination assay. Linear ubiquitin formation assays were performed in 50 mM HEPES pH 7.9, 100 mM NaCl, 10 mM MgCl₂ and 0.6 mM DTT using 200 nM E1, 1 μ M UbcH5B, 1 μ M HOIP RBR or HOIP RING2L and 40 μ M untagged ubiquitin. Reactions were started by addition of 10 mM ATP and were incubated at 30 °C for 2 h. Samples were taken at the indicated time points and treated with 50 mM sodium acetate pH 4.5 as described previously⁶, mixed with SDS sample buffer and analysed by SDS-PAGE using 12% Bolt Bis-Tris gels (Life Technologies). Proteins were visualized with Coomassie Brilliant blue dye. To test the activating effect of linear di-ubiquitin on auto-inhibited HOIP UBA-RBR, 5 μ M HOIP UBA-RBR was pre-incubated with N- and C-terminally blocked linear di-ubiquitin or HOIL-1L at the indicated concentrations before addition of the remaining assay components. Samples were taken after 60 min and subsequently treated as described above.

UbcH5B~ubiquitin to HOIP RBR ubiquitin transfer assay. To monitor HOIP~ubiquitin thioester ubiquitin transfer from UbcH5B to HOIP, Ube1 (100 nM), UbcH5B (4 μ M) and N-terminally blocked ubiquitin (32 μ M) were mixed in 50 mM HEPES pH 7.9, 100 mM NaCl, 10 mM MgCl₂ and 5 mM ATP and incubated at 25 °C for 5 min when 2 μ M HOIP RBR was added. Samples were taken 10 s after HOIP addition, quenched by addition of pre-heated SDS protein-loading buffer without DTT, and run on a 12% SDS-PAGE gel (Life Technologies). The 10-s time point used was empirically determined with a time-course experiment (Extended Data Fig. 9g). Gels were stained with Coomassie Brilliant blue dye and scanned on a Li-COR Odyssey scanner using the 700 nm (red) channel. For the thioester transfer assay shown in Fig. 3d, 200 nM Ube1, 2 μ M UbcH5B, 8 μ M HOIP RBR, 8 μ M N-terminally blocked ubiquitin and 10 mM ATP were used and samples

taken after 30 s. Furthermore, proteins were transferred to a PVDF membrane and ubiquitin was visualized on a Li-COR Odyssey scanner at 800 nm using an anti-ubiquitin antibody (P4D1, Santa Cruz, 1:200 dilution in TBST (50 mM Tris pH 7.4, 150 mM NaCl, 0.05% Tween-20)) followed by an IRDye 800CW secondary antibody (LI-COR, 1:10,000 dilution in TBST). All quantitative experiments shown in graphs were performed in triplicates and band intensities were quantified using the ImageStudio software (LI-COR). HOIP thioester transfer activity was calculated as the fraction of HOIP~ubiquitin to total HOIP for each mutant and normalized against thioester transfer activity of wild-type HOIP. Data were analysed in GraphPad Prism using two-tailed unpaired Student's *t*-test or one-way ANOVA followed by Tukey's post hoc test.

Allosteric activation of HOIP RBR by di-ubiquitin. To test the allosteric activation of HOIP RBR by linear di-ubiquitin, a modified ubiquitin transfer assay was performed. HOIP RBR was pre-incubated with N- and C-terminally blocked linear di-ubiquitin at the indicated final concentrations for 5 min at 25 °C. At the same time, Ube1, UbcH5B, ubiquitin and ATP were premixed and incubated for 5 min at 25 °C, resulting in fully loaded UbcH5B~ubiquitin. Both mixtures were subsequently mixed together, resulting in final concentrations of 100 nM Ube1, 2 μ M UbcH5B, 8 μ M N-terminally blocked ubiquitin and 2 μ M HOIP RBR in the final buffer described for the standard ubiquitin transfer assay. Samples were taken after 30 s and further treated as described for the standard transfer assay. A 30-s time point was determined to give the best results in this assay, in which lower E2 and mono-ubiquitin concentrations were used, resulting in an overall slower reaction rate. The experiments comparing the effects of linear versus K48-linked di-ubiquitin (Extended Data Fig. 9e) were performed similarly, with the difference that all samples were incubated with apyrase (Sigma) for 5 min to deplete ATP before addition of HOIP/di-ubiquitin and prevent E2-loading of K48-linked di-ubiquitin, which features a free C terminus on one of the ubiquitin units.

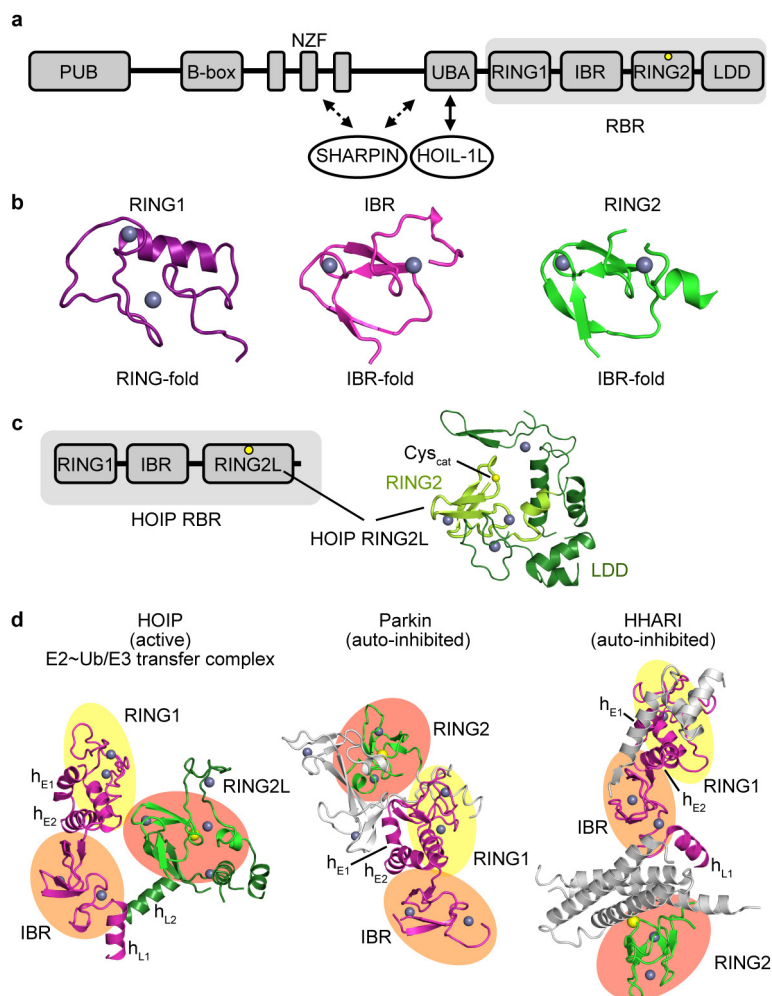
Analytical ultracentrifugation (AUC). Sedimentation equilibrium experiments were performed in a ProteomeLab XL-I (Beckman Coulter) analytical ultracentrifuge. HOIP RBR/UbcH5B~ubiquitin as used for crystallization was loaded into a 6-channel equilibrium cell at 5.0, 2.5 and 1.25 μ M concentration and centrifuged at 10,000 r.p.m., 20 °C in an An-50 Ti 8-place rotor until equilibrium was achieved. Data were analysed using HeteroAnalysis software (J. L. Cole and J. W. Lary, University of Connecticut; <http://www.biotech.uconn.edu/auf/>).

Isothermal titration calorimetry (ITC). ITC experiments were performed on an ITC200 calorimeter (Microcal). Aliquots (2 μ l each) of 500–650 μ M UbcH5B~ubiquitin or di-ubiquitin solution were injected into the cell containing 40–50 μ M HOIP RBR or HOIP RBR/di-ubiquitin complexes. The experiments were performed at 23 °C in buffer containing 10 mM HEPES pH 7.9, 100 mM NaCl. For titrations of UbcH5B~ubiquitin into HOIP RBR/di-ubiquitin complexes, HOIP RBR was pre-incubated with an equimolar amount of di-ubiquitin before the ITC experiments. Data were analysed using the Origin software (Microcal).

NF- κ B luciferase assay. Human embryonic kidney (HEK) 293T cells (ATCC) were co-transfected with NF- κ B-luc reporter plasmid that contains an NF- κ B response element upstream of the promoter driving the luciferase reporter gene, pGL4.74[hRluc/TK] control vector (Promega) and epitope tagged Flag-HOIP or myc-HOIL-1L pcDNA3.1(+) plasmids in 6-well plates in triplicates using Lipofectamine 2000 transfection reagent. Since this assay could be carried out in a variety of cellular contexts, HEK293T cells were used because they are easy to transfect and suitable for the assay. The cells tested negative for mycoplasma contamination. Empty pcDNA3.1(+) vector was used as control. After 36 h, cells were lysed and 20 μ l cell lysates were used to measure firefly luciferase and Renilla luciferase (transfection control) signals using the dual luciferase reporter assay system according to the manufacturer's protocol (Promega). Data were analysed in GraphPad Prism and one-way ANOVA followed by Tukey's post hoc tests were used for statistical analysis. Immunoblotting was performed with anti-Flag (clone M2, Sigma-Aldrich) and anti-myc (clone 9E10, Sigma-Aldrich) antibodies, to confirm equivalent wild-type and mutant protein expression levels.

- Luna-Vargas, M. P. *et al.* Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. *J. Struct. Biol.* **175**, 113–119 (2011).
- Brzovic, P. S., Lissounov, A., Christensen, D. E., Hoyt, D. W. & Kleit, R. E. A. UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Mol. Cell* **21**, 873–880 (2006).
- Hilgart, M. C. *et al.* Automated sample-scanning methods for radiation damage mitigation and diffraction-based centering of macromolecular crystals. *J. Synchrotron Radiat.* **18**, 717–722 (2011).
- Stepanov, S. *et al.* JBLuice-EPICS control system for macromolecular crystallography. *Acta Crystallogr. D* **67**, 176–188 (2011).
- Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).

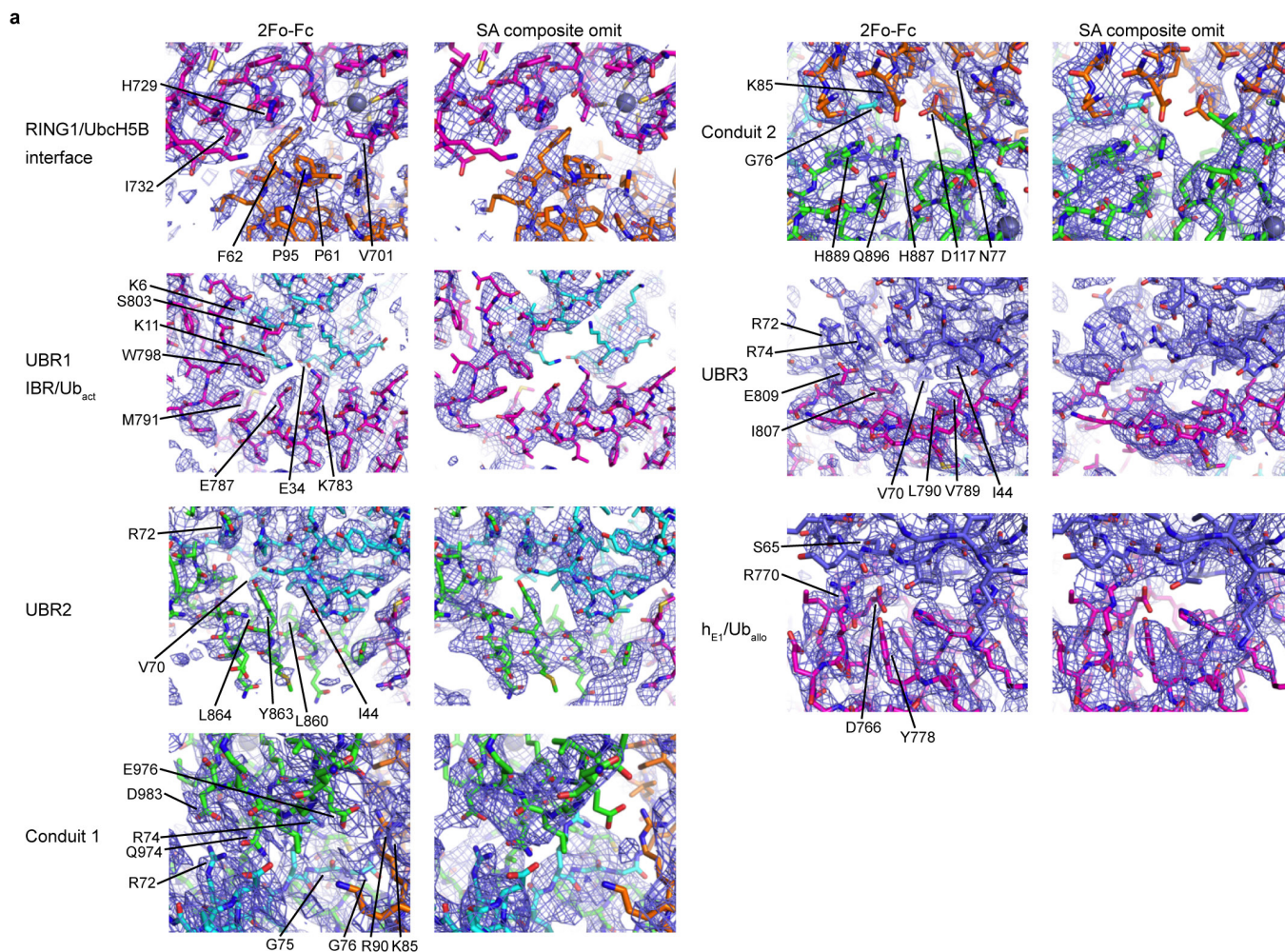
36. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
37. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
39. Sakata, E. *et al.* Crystal structure of UbcH5b~ubiquitin intermediate: insight into the formation of the self-assembled E2~Ub conjugates. *Structure* **18**, 138–147 (2010).
40. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
41. Bunkóczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D* **67**, 303–312 (2011).
42. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
43. Nicholls, R. A., Long, F. & Murshudov, G. N. Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. D* **68**, 404–417 (2012).
44. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).
45. Ozkan, E., Yu, H. & Deisenhofer, J. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl Acad. Sci. USA* **102**, 18890–18895 (2005).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
47. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62**, 439–450 (2006).
48. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCr* **1**, 213–220 (2014).
49. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
50. Dong, K. C. *et al.* Preparation of distinct ubiquitin chain reagents of high purity and yield. *Structure* **19**, 1053–1063 (2011).
51. Elliott, P. R. *et al.* Molecular basis and regulation of OTULIN-LUBAC interaction. *Mol. Cell* **54**, 335–348 (2014).
52. Spratt, D. E. *et al.* A molecular explanation for the recessive nature of parkin-linked Parkinson's disease. *Nature Commun.* **4**, 1983 (2013).
53. Spratt, D. E., Mercier, P. & Shaw, G. S. Structure of the HHAR catalytic domain shows glimpses of a HECT E3 ligase. *PLoS ONE* **8**, e74047 (2013).
54. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
55. Grishin, A. M. *et al.* Structural basis for the inhibition of host protein ubiquitination by Shigella effector kinase OspG. *Structure* **22**, 878–888 (2014).
56. Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Principles of Biochemistry*. 2nd edn, (Worth Publishers, 1993).
57. Afonine, P. V. *et al.* FEM: feature-enhanced map. *Acta Crystallogr. D* **71**, 646–666 (2015).
58. Emmerich, C. H. *et al.* Activation of the canonical IKK complex by K63/M1-linked hybrid ubiquitin chains. *Proc. Natl Acad. Sci. USA* **110**, 15247–15252 (2013).
59. Sato, Y. *et al.* Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of the HOIL-1L subunit of the linear ubiquitin chain assembly complex. *Proc. Natl Acad. Sci. USA* **108**, 20520–20525 (2011).



Extended Data Figure 1 | HOIP domain organization and nomenclature.

a, Domain organization of HOIP as commonly outlined in the literature. HOIP consists of a PNGase/ubiquitin-associated (PUB) domain followed by a B-box zinc-finger (B-box) domain⁵¹, NPL4 zinc-fingers (NZF), the auto-inhibitory UBA domain and the RING-between-RING module (RBR, grey background). The HOIP RBR module contains the typical RING1, in-between RING (IBR) and RING2 domains, and a HOIP-specific additional linear ubiquitin chain determining domain (LDD). A yellow circle indicates the RBR catalytic cysteine (C885) forming the HECT-like thioester intermediate with ubiquitin. The binding sites of the other LUBAC constituents HOIL-1L and SHARPIN are also indicated. **b**, The RBR RING2 domain has the topology of an IBR domain. The individual HOIP RING1, IBR and RING2 domains from the HOIP RBR/E2~Ub/Ub structure are shown to enable direct comparison of their folds. This illustrates that the zinc-finger domain designated RING2 in fact adopts the topology of an IBR, as multiple groups have reported for various RBR E3 ligases previously^{8–11,14,52,53}. The terms RBR and RING2 however are used in this study for consistency with the widely accepted nomenclature. **c**, The HOIP RING2–LDD region. HOIP features

an extension of its catalytic RING2 domain termed LDD, which adds two zinc-fingers and a helical arrangement to the RING2. The LDD is usually denoted as a domain following RING2. However, Rittinger and colleagues¹⁴ showed that the LDD is intertwined with the HOIP RING2 to form a single extended domain that contains a central canonical RBR RING2 with the additional features of the LDD ensuring the linear ubiquitin chain formation characteristic of HOIP. This domain will thus be designated RING2L (for RING2–LDD). The RING2L from the current HOIP structure is displayed with RING2 in light green and LDD in dark green. **d**, The structural arrangement of active HOIP RBR in the HOIP/E2~ubiquitin complex is markedly different from that of auto-inhibited RBRs. Left, active HOIP RBR from the HOIP/E2~ubiquitin complex. The RING1–IBR region and the RING2L are coloured magenta and green respectively. The individual RBR domains are also highlighted: RING1, yellow circle; IBR, orange circle; RING2, red circle. The RING1 extension helices (h_{E1}, h_{E2}) and IBR–RING2 linker helices (h_{L1} and h_{L2}) are labelled. Middle and right, analogous representations of auto-inhibited Parkin and HHARI (PDB: 411H (ref. 8) and PDB: 4KBL (ref. 11)). Additional domains and regions besides the RBR of Parkin and HHARI are coloured grey.



b

Data collection and refinement statistics

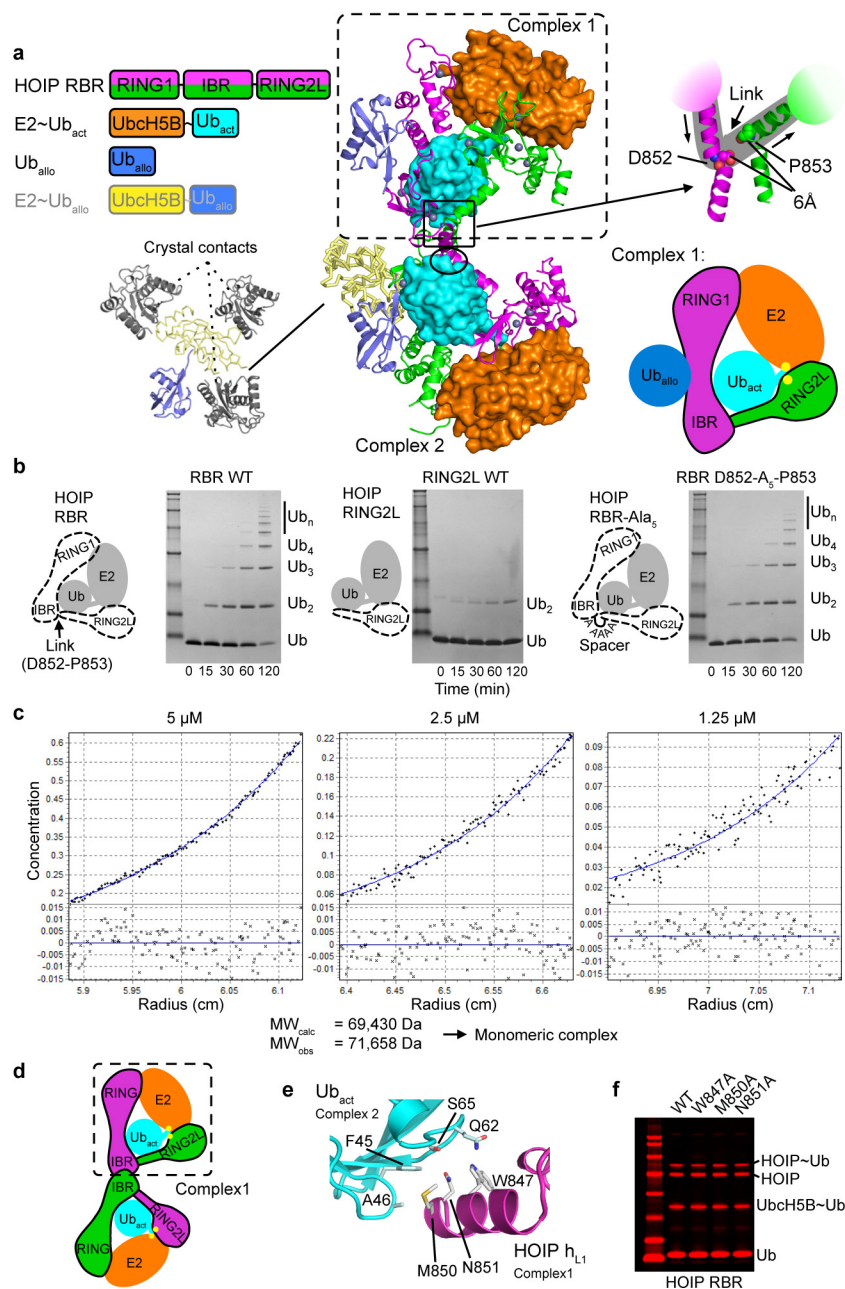
	HOIP RBR/Ubch5B~Ub/Ub
Data collection	
Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	104.82, 75.74, 120.96
α , β , γ (°)	90, 95.56, 90
Resolution (Å)	29.69–3.48 (3.72–3.48)*
<i>R</i> _{merge}	0.093 (0.875)
$\langle I/\sigma \rangle$	9.9 (1.6)
CC1/2	0.995 (0.648)
Completeness (%)	98.6 (96.1)
Redundancy	6.5 (6.2)
Refinement	
Resolution (Å)	29.69–3.48
No. reflections	22930
<i>R</i> _{work} / <i>R</i> _{free}	24.87 / 30.26
No. atoms	
Protein	10953
Ligand/ion	16
Water	0
B-factors	
Protein	185.8
Ligand/ion	156.6
Water	---
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.465

Data was collected from a single crystal.

*Highest resolution shell is shown in parenthesis.

Extended Data Figure 2 | Quality of crystallographic data and electron density maps. a, Final $2F_o - F_c$ (left) and simulated annealing (SA) composite omit (right) electron density maps of select interfaces of the

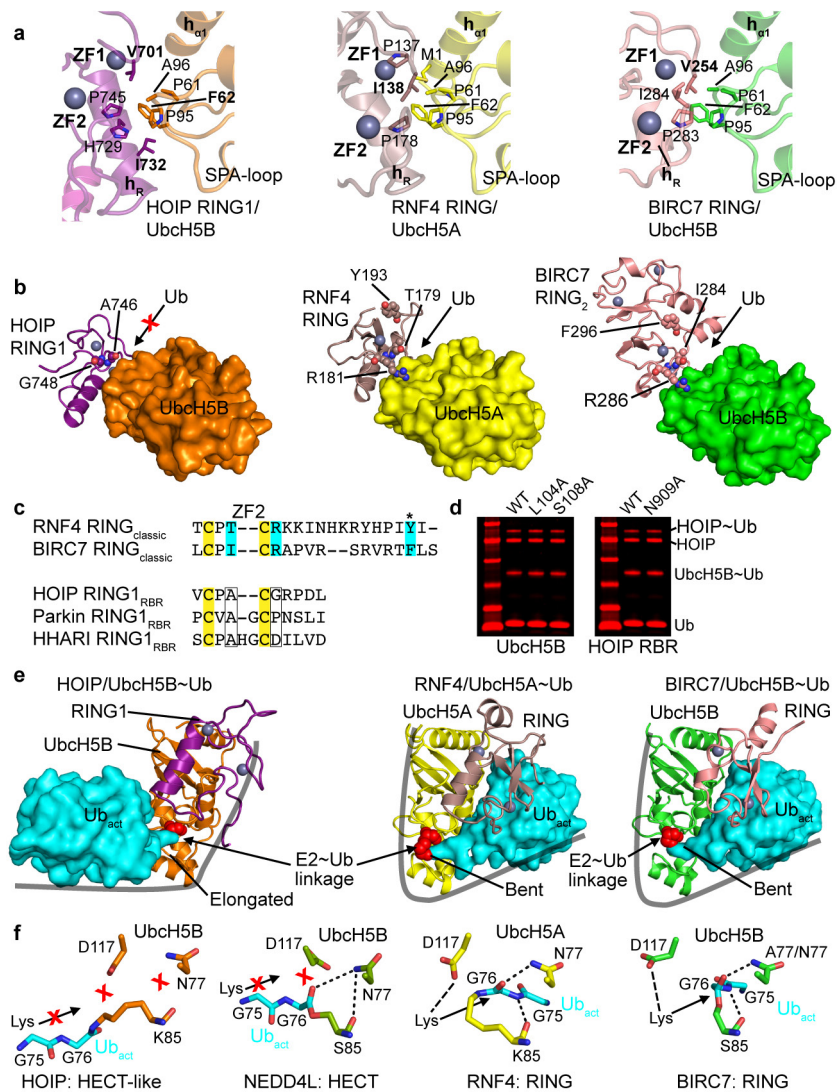
HOIP/Ubch5B~Ub/Ub complex contoured at 1σ . Proteins are shown in sticks and coloured according to Fig. 1. **b**, Data collection and refinement statistics.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Complexity of the crystallographic asymmetric unit and structure of the HOIP/UbcH5B~ubiquitin/ubiquitin E2-E3 transfer complex. **a**, The asymmetric unit contains two transfer complexes. Left, colour schematic of the proteins present in the asymmetric unit. Middle, structure of the asymmetric unit. The asymmetric unit contains two HOIP RBR/UbcH5B~ubiquitin complex arrangements (complex 1, 2). The two UbcH5B~ubiquitin conjugates are coloured orange and cyan, respectively, and are bound to two HOIP RBR molecules (magenta and green), which cross over between the complexes. Additional allosteric ubiquitin (Ub_{allo}, blue) and UbcH5B~ubiquitin (UbcH5B~Ub_{allo}, yellow and blue, respectively) molecules are bound to the HOIP RBRs in complex 1 and complex 2, respectively. Since Ub_{allo} makes all contacts with the RBRs and the UbcH5B of the UbcH5B~Ub_{allo} conjugate solely mediates crystal contacts (bottom left), only Ub_{allo} of complex 1 is displayed and discussed in the text and figures in terms of the additional ubiquitin binding. The black oval indicates an additional HOIP h_{L1}/Ub_{act} inter-complex interaction discussed in panels **e** and **f**. Right, close-up of the region where the two RBRs of HOIP cross over between the complexes. The close-up shows that residues D852 and P853 of the respective RBRs come in 6 Å proximity suggesting a continuity in the biological complex in which the two residues from the respective RBRs are linked (as indicated by the grey background), resulting in the monomeric complex schematically illustrated underneath and discussed in panels **c–f**. **b**, The RING1–IBR and RING2L form two distinct entities to bind E2~ubiquitin. The monomeric complex as displayed for complex 1 assumes a flexible linkage between the autonomous units of the RING1–IBR arm and the RING2L from the two different RBR molecules in the asymmetric unit. This linkage is formed by residues D852 and P853 connecting IBR and RING2L (schematically illustrated in the cartoon). To test the structural integrity of the assumed link and the autonomy of the RING1–IBR arm on one side and the RING2L on the other side, we introduced a spacer comprising five alanine residues between D852 and P853 in HOIP RBR (see cartoon) and measured the activity of the RBR D852–Ala₅–P853 insertion mutant in polyubiquitination assays. The assays show that the mutant (right) retains an activity similar to the wild-type RBR (left) indicating that indeed RING1–IBR and RING2L act as autonomous units. The dramatically reduced activity of HOIP RING2L alone (residues P853 to end) is also shown for reference (middle). **c**, The HOIP RBR/UbcH5B~ubiquitin complex is monomeric at concentrations

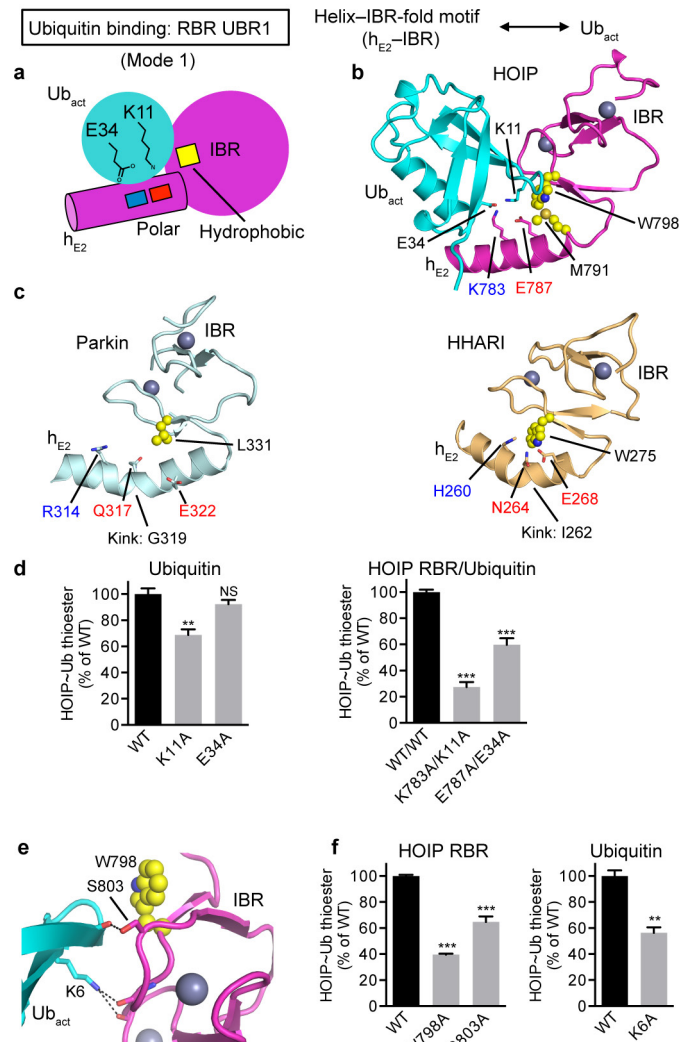
of 1.25–5 µM. To determine if the HOIP RBR/UbcH5B~ubiquitin complex is indeed monomeric in solution, we analysed the isolated HOIP RBR/UbcH5B~ubiquitin complex protein material that was used for crystallization by sedimentation equilibrium analytical ultracentrifugation (SE-AUC). SE-AUC provides an absolute, shape-independent measurement of molecular weight, thus allowing accurate determination of the oligomeric state. The three SE-AUC experiments performed on the HOIP RBR/UbcH5B~ubiquitin complex yielded an absolute molecular weight (MW) of 71,658 Da, indicating a monomeric complex. At an order of magnitude higher concentrations (12.5–50 µM), SE-AUC results indicate the formation of a dimer with a MW of ~144 kDa, although curve fitting residuals also show substantial presence of aggregates (data not shown). These results indicate that the biological complex in solution is monomeric at physiological low µM concentrations such as those used for the thioester transfer assays and polyubiquitination assays. However, the dimeric arrangement observed in the crystal structure might be relevant in a high concentration setting such as within the LUBAC complex. Here, a high local concentration of HOIP RBR could favour binding of the E2~ubiquitin between the RING1–IBR and RING2L of two neighbouring molecules. Importantly, all mechanisms depicted in this article hold true for both the monomeric and dimeric states (as illustrated in **d**). This means that the deduced mechanism is in principle applicable to different RBR E3 ligases of which some might function as dimers in local high concentration assemblies (such as within the LUBAC), whereas others might be active in a monomeric setting. **d**, Schematic illustration of the dimeric arrangement as observed in the asymmetric unit. The schematic shows that all features deduced (Figs 1–4 and Extended Data Figs 4–10) are also valid for the dimeric case (binding of Ub_{allo} is omitted for clarity). **e**, Asymmetric unit dimer-related interactions between HOIP h_{L1} and Ub_{act}. The dimeric arrangement contains no additional protein–protein interfaces compared to the monomeric assemblies with the exception of h_{L1} residues W847, M850 and N851, which in the asymmetric unit contact the activated ubiquitin of the other complex (indicated by an oval in **a**). **f**, Mutational analysis of HOIP h_{L1}/Ub_{act} interactions. Mutations of HOIP h_{L1} residues that interact with Ub_{act} have no effect on thioester transfer activity (Coomassie-stained bands in red), indicating that this ‘trans’ complex interaction is not critical for the RBR mechanism, in line with the model of a monomeric arrangement.



Extended Data Figure 4 | See next page for caption.

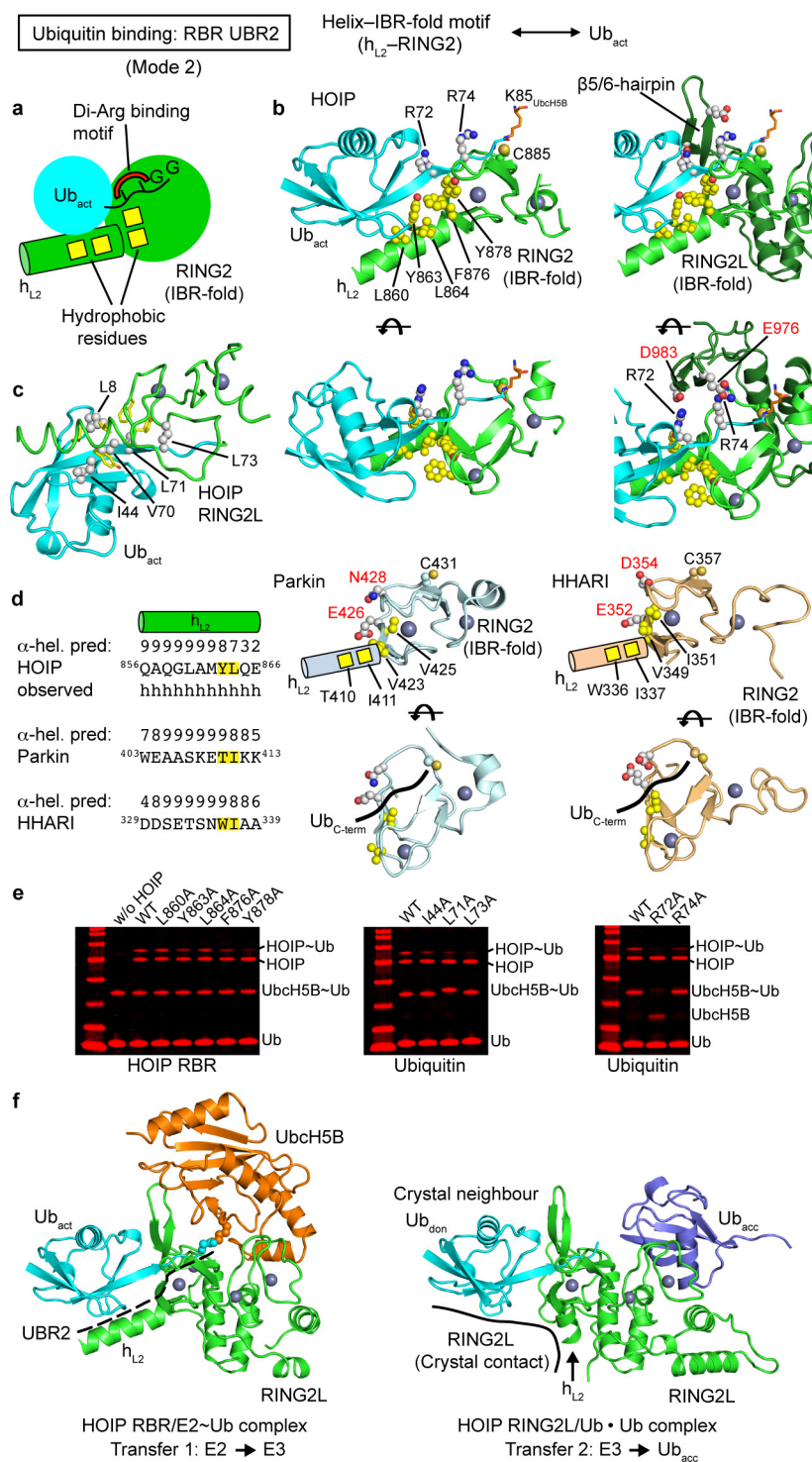
Extended Data Figure 4 | The HOIP RING1-IBR applies an altered binding mode compared to classic RING E3s necessitating a HECT-like mechanism. **a**, The HOIP RBR RING1 uses an E2 interaction pattern similar to classic RINGs, but which results in a shifted binding. Shown are the details of the RING/E2 interaction in the HOIP RING1_{RBR}/UbcH5B~ubiquitin complex (left), the RNF4 RING_{classic}/UbcH5A~ubiquitin complex (middle; PDB: 4AP4 (ref. 21)), and the BIRC7 RING_{classic}/UbcH5B~ubiquitin complex (right; PDB: 4AUQ (ref. 24)). The HOIP RBR-type RING1 uses a pattern of hydrophobic residues as the core of the interaction with E2 that is similar to that in classic RING E3 ligases. Subtle differences however support a shifted binding mode (see also Fig. 2b). The main features of the RING and E2 as well as HOIP residues mutated in Fig. 2e–g are displayed in bold. Zinc-finger (ZF) 1 and 2 of the RING domains and the SPA-loop of the E2 containing a conserved Ser-Pro-Ala motif are annotated. For the following panels the same structures and colour codes as in **a** are used. **b**, The shift in binding and altered surface residues in HOIP RING1 do not support the composite RING/E2 binding site for activated ubiquitin used by classic RING/E2 complexes. UbcH5A/B E2s are rendered as surface representation and the RING domains in ribbon representation. Residues crucial for classic RING E3s to recruit the activated ubiquitin in the composite RING/E2 ubiquitin-binding surface^{21,24} are depicted (middle, right). In HOIP RING1 (left) equivalent residues are not conserved (displayed), indicating that HOIP cannot accommodate the activated ubiquitin in its RING1/UbcH5B complex. For illustration purposes, only one monomer of the dimeric RNF4 is shown (although Y193 from the other RING molecule is still displayed)²¹; for BIRC7 the RING dimer (RING₂) is displayed²⁴. **c**, Alignment of HOIP, Parkin and HHARI RING1 domains with classic RING domains centred around residues displayed in **b**. Residues crucial for ubiquitin binding in classic RINGs are highlighted in cyan and their structural equivalents in the RING1 domains of the RBR E3 ligases HOIP, Parkin and HHARI are indicated by boxes, attesting to the absence of a composite RING1/E2 ubiquitin-binding site in RBR ligases (asterisk indicates ubiquitin interaction residues from the other RING molecule in the dimers formed by the classic RINGs RNF4 and BIRC7). The T/I-C-R sequence observed in the dimeric RINGs of RNF4 and BIRC7 represents the highly conserved Φ -x-R/K motif, where Φ is a hydrophobic residue and x is either a Cys in RING E3 ligases or a polar residue in U-box ligases²⁶. This motif is not only critical for E3-mediated catalysis by dimeric RING ligases (such as RNF4 or BIRC7) but is also necessary for E2-mediated catalysis by simpler monomeric RING and U-Box E3 ligases²⁶. The fact that this motif is not conserved in HOIP, HHARI and Parkin further confirms the mechanistic differences between RBRs

and classic RING domains. **d**, Thioester transfer assays show that E2 residues critical for classic RING-supported catalysis are not important for HOIP catalysis. Left, HOIP RBR thioester transfer assays show similar activity of wild-type UbcH5B and L104A and S108A mutants. This is in stark contrast to the reported effects of these mutations on classic RING-supported catalysis^{21,24}, underlying the fundamentally different mechanism of the HECT-like catalysis by HOIP RBR. Right, mutation to Ala of HOIP RBR N909, which would be in the vicinity of the activated ubiquitin if the E2~ubiquitin conjugate were bound in a bent manner (see **e**), also shows no effect on HOIP thioester formation (Coomassie-stained bands in red). **e**, The altered E2~ubiquitin binding mode of RBR RING1 results in the requirement for a HECT-like mechanism. Displayed are the entire RING/UbcH5~ubiquitin complexes with RINGs and E2s depicted in ribbon representation and the activated ubiquitin in surface representation. The bipartite binding mode used by the HOIP RING1-h_{E2}-IBR arm (see also Fig. 2a) results in an elongated E2~ubiquitin conformation (left, only the RING1 domain of HOIP is depicted) while formation of a composite RING/E2 binding surface in the case of classic RING E3 ligases (middle, right) results in binding of the activated ubiquitin in a compact manner with a bent E2~ubiquitin conformation. Importantly, this bent conformation places the thioester link in a specific position relative to the catalytic machinery of the E2, allowing direct attack by the lysine/amine function of a substrate or growing ubiquitin chain. The Lys85/Ser85 residues mediating the E2~ubiquitin linkage and mimicking UbcH5A/B catalytic cysteine C85 are displayed as red spheres. In the elongated E2~ubiquitin conformation propagated by the HOIP RBR, this attack is not possible. The linkage is however ideally positioned for the attack by the RBR catalytic cysteine in a HECT-like mechanism (see also Fig. 3 and Extended Data Fig. 7a). **f**, Close-up of the catalytic centres in E2~ubiquitin linkages. Details of the catalytic centres resulting from the E2~ubiquitin conjugate conformations outlined in **e** and, for comparison, the HECT-type E3 NEDD4L/UbcH5B~Ub structure (PDB: 3JW0 (ref. 28)), with the directionality of an attacking amine indicated as previously proposed²¹. In the HECT-like RBR arrangement, the UbcH5B~ubiquitin linkage is not aligned correctly relative to the E2 catalytic machinery for a direct attack by an amine function. This is similar in the HECT-type arrangement in the NEDD4L complex but completely different from the arrangement in classic RING-supported E2 catalysis. Additionally, the ubiquitin C-terminal residues G75–G76 reside in a position that would overlap with the attacking amine. The available structure of the BIRC7/UbcH5B~Ub complex (PDB: 4AUQ (ref. 24)) features an UbcH5B N77A mutant and the remainder of the Asn side-chain has been manually added based on wild-type UbcH5B from PDB: 2ESK (ref. 45) (right).



Extended Data Figure 5 | Helix-IBR-fold motifs constitute new ubiquitin-binding regions (UBR) in active RBR proteins: binding of the activated ubiquitin by UBR1 using binding mode 1. **a**, Schematic illustrating binding mode 1 used by h_{E2} -IBR to bind the activated ubiquitin in the RING1-IBR arm. The general principle of this binding mode is that the RING1 extension helix 2 (h_{E2}) preceding the IBR presents a pattern of charged/polar residues (indicated by blue and red squares, which symbolize K, R, H and E, Q, N residues respectively) that interact with ubiquitin E34 and K11. These interactions are supported by the IBR surface, with a particular contribution of hydrophobic residues (yellow square) flanking the salt bridge system. **b**, Coordination of the activated ubiquitin by HOIP h_{E2} -IBR in mode 1. HOIP h_{E2} residues K783 and E787 bind ubiquitin residues E34 and K11 and are flanked by hydrophobic residues M791 and W798 from HOIP IBR. **c**, Structurally equivalent residues in Parkin and HHARI. Displayed are the h_{E2} -IBR modules from auto-inhibited Parkin (PDB: 5C1Z (ref. 12)) and HHARI (PDB: 4KBL (ref. 11)) with residues equivalent to HOIP residues in **b** depicted, illustrating the general conservation of UBR1. It should be noted that these structures feature auto-inhibited forms of the RBR proteins, which exhibit a kink in h_{E2} of UBR1. This kink would sterically hinder ubiquitin binding to UBR1 and probably participates in the RBR auto-inhibition

mechanism (see also Extended Data Fig. 8). **d**, Thioester-transfer assays of ubiquitin and UBR1 salt bridge mutants. In agreement with the observed four-residue salt bridge system in **b**, the single K11A or E34A ubiquitin mutations show only a slight to moderate effect since the remaining charged residue can still coordinate the two oppositely charged residues of HOIP. In contrast, elimination of both similarly charged residues in the complex by combining the HOIP K783A and ubiquitin K11A or HOIP E787A and ubiquitin E34A mutations results in a more dramatic loss of activity (mean activity \pm s.e.m. ($n=3$), one-way ANOVA followed by Tukey's post hoc test; $**P < 0.01$; $***P < 0.001$; NS, not significant; representative gels shown in Supplementary Fig. 1). **e**, **f**, Role of the IBR in UBR1. Close-up of the additional IBR/ Ub_{act} interactions in stick representation (**e**) shows that HOIP S803 and ubiquitin K6 coordinate the backbone carbonyl functions of ubiquitin T12 and HOIP A800/K829, respectively. W798, which is involved in hydrophobic interactions, is also displayed in sphere representation. Quantitative thioester transfer assays (**f**) show that alanine mutants of residues outlined in **e** cause a significant loss of activity (mean activity \pm s.e.m. ($n=3$), left: one-way ANOVA followed by Tukey's post hoc test, right: two-tailed unpaired Student's t -test; $**P < 0.01$; $***P < 0.001$; representative gels shown in Supplementary Fig. 1).

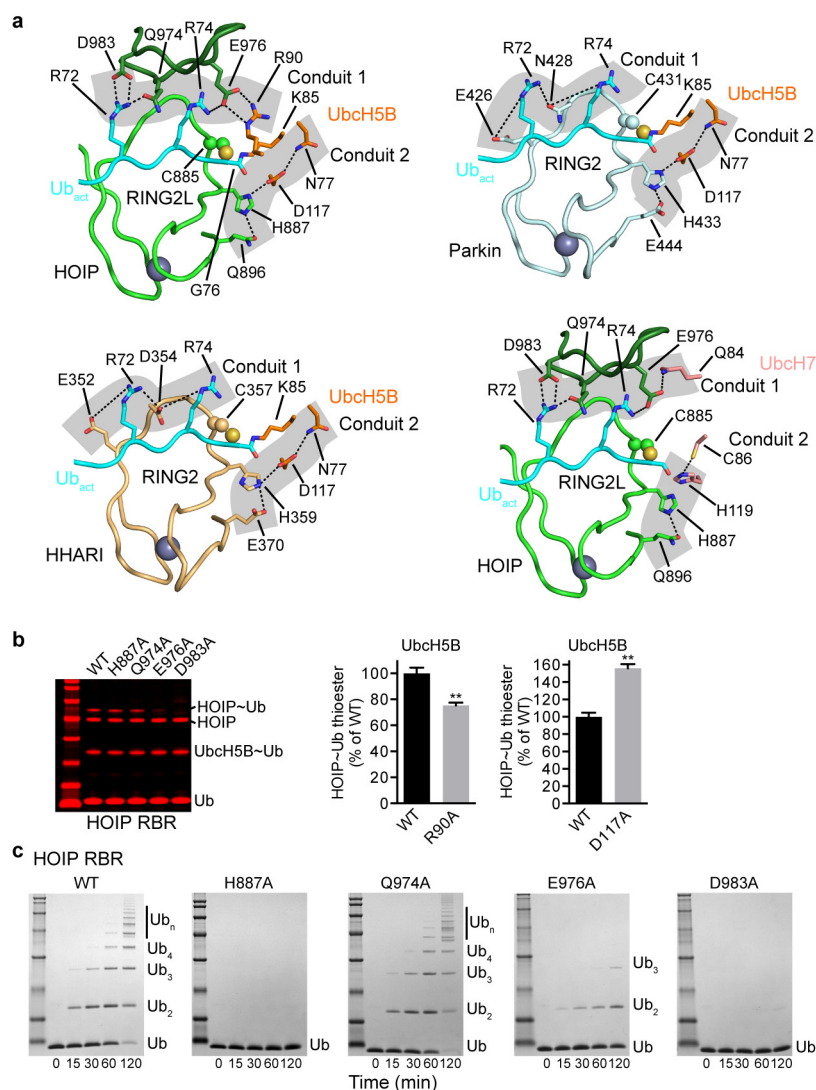


Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Binding of the activated ubiquitin by UBR2 using binding mode 2 and exclusive binding of E2 and acceptor ubiquitin.

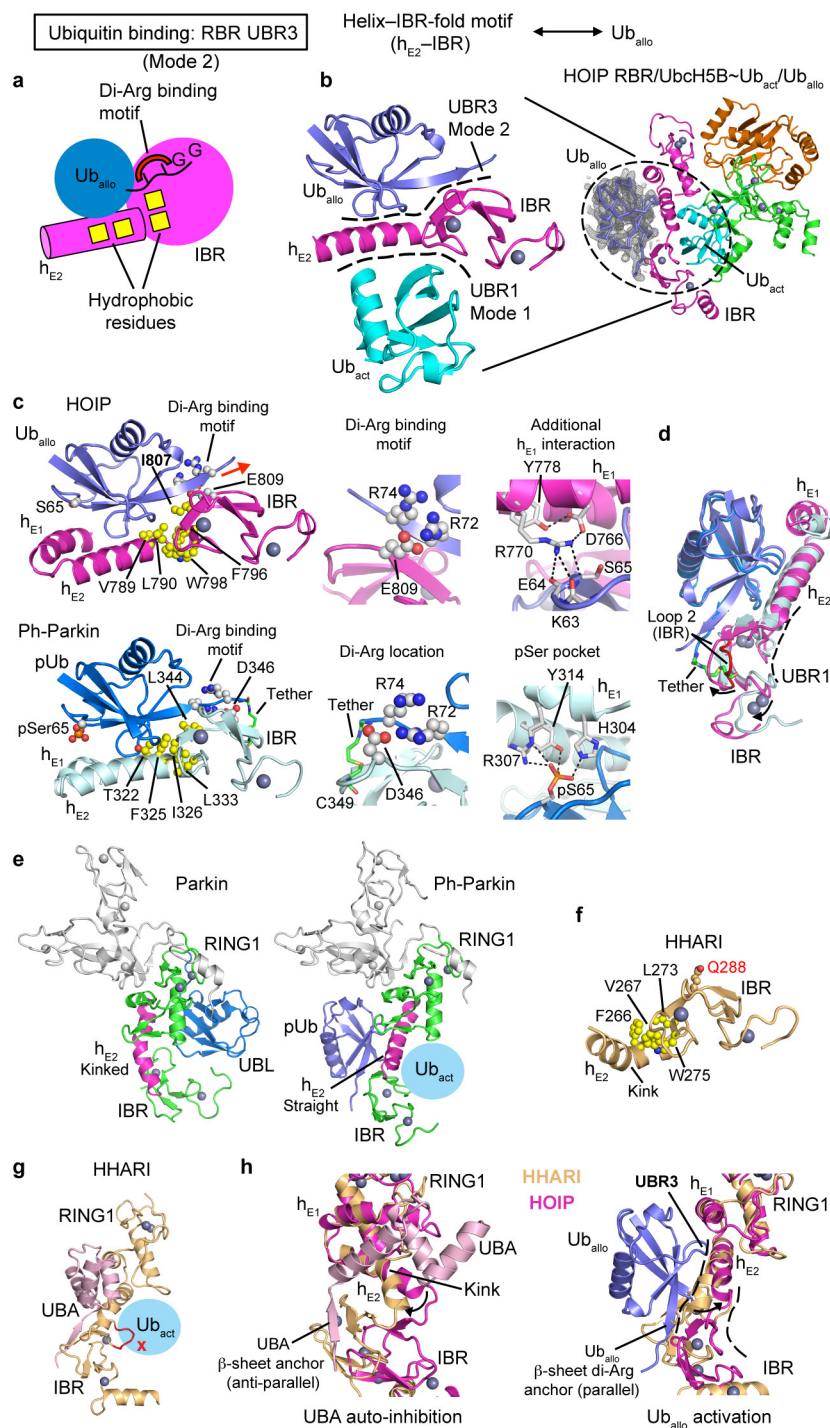
a, Schematic illustrating binding mode 2 used by a helix–IBR-fold motif ($h_{1,2}$ –RING2) to bind the activated ubiquitin and position the thioester linkage for the transfer reaction. The second helix ($h_{1,2}$) of the linker between the IBR domain and the catalytic RING2 domain uses a pattern of two or three hydrophobic residues (yellow squares) to interact with the ubiquitin canonical hydrophobic patch surrounding I44 (ref. 28; not shown). Hydrophobic residues of the RING2 IBR-fold complete the hydrophobic interaction network by coordinating residues L71 and L73 in the second hydrophobic patch²⁸ of ubiquitin (not shown). The central hallmark of this binding mode is the coordination of the characteristic di-Arg (R72, R74) motif in the ubiquitin C terminus, resulting in a firm placement of the C terminus on ZF1 of RING2. **b**, Structure of the interaction of the helix–IBR-fold in HOIP $h_{1,2}$ –RING2 (UBR2) with the activated ubiquitin. Left, $h_{1,2}$ residues L860, Y863 and L864 (yellow spheres) interact with ubiquitin residues L8, I44 and V70 (not shown). Additionally, hydrophobic residues F876 and Y878 (yellow spheres) from the IBR-fold of the minimal catalytic RING2 (light green, see also Extended Data Fig. 1c) coordinate ubiquitin residues L8, L71 and L73 (not shown). Right, display of the full HOIP RING2 including the LDD insertion (RING2L). The coordination of the ubiquitin di-Arg motif is achieved by HOIP residues D983 and E976 from the LDD insertion that is part of the catalytic HOIP RING2L (dark green). This results in the placement of the E2~Ub thioester linkage (K85 replacing UbCH5B C85 is shown as orange sticks) in the vicinity of the catalytic HOIP C885. Bottom, alternatively oriented views of the interaction. **c**, The two hydrophobic patches of ubiquitin engaged by UBR2. The hydrophobic residues of HOIP RING2 interacting with ubiquitin as highlighted in **b** are shown as yellow sticks. The interaction residues on the canonical hydrophobic patch of ubiquitin (L8, I44, V70) and the second hydrophobic patch (L71, L73)²⁸ are displayed as grey spheres. **d**, The RING2 domains of Parkin and HHARI also contain a helix–IBR-fold ($h_{1,2}$ –RING2) module with patterns of residues consistent with the formation of a UBR2. Left, helical predictions for the region preceding the RING2 domains of HOIP, Parkin and HHARI. The structures of Parkin and HHARI in their auto-inhibited forms do not display a helix equivalent to $h_{1,2}$ because this region is either not defined (in the crystal structures of HHARI and most Parkin structures) or adopts an extended conformation (in two other Parkin structures)^{8–11}. However, a helical prediction reliability score (with 1 lowest to 9 highest score) calculated using JPred4⁵⁴ shows a strong helical probability for the segment of Parkin and HHARI preceding the RING2 domain. In fact, the score is similar to that of HOIP, which is displayed with the observed helical secondary structure, pointing to the presence of an equivalent of $h_{1,2}$ in active forms of Parkin and HHARI.

These RBR E3 ligases also contain residues capable of interacting with the hydrophobic patch in ubiquitin in positions equivalent to HOIP Y863 and L864 (highlighted in yellow). Right, structures of the RING2 domains of PARKIN (PDB: 4I1H (ref. 8)) and HHARI (PDB: 4KBL (ref. 11)) showing hydrophobic residues (yellow sphere representation) in structurally equivalent positions to HOIP F876 and Y878 and residues (labelled red) capable of interacting with the di-Arg motif in their catalytic RING2. Helix $h_{1,2}$ with the conserved hydrophobic residues not present in the crystal structures as discussed above is indicated schematically. Bottom, different orientations with the putative placement of the ubiquitin C terminus indicated schematically. **e**, The effect of UBR2 alanine mutations in thioester transfer assays increases with their proximity to the di-Arg motif. Left, mutation of HOIP UBR2 hydrophobic residues to alanine. The $h_{1,2}$ L860A and L864A mutations show little effect on activity, while the Y863A mutation and particularly the RING2L F876 and Y878 mutations, which reside proximal to the di-Arg binding motif formed by D983 and E976 (see also Extended Data Fig. 7) show a marked reduction in activity. Middle, right, mutation of complementary ubiquitin residues involved in UBR2 binding. Similarly to the HOIP mutations, the ubiquitin I44A, L71A and L73A mutations show increasing effects with a closer location to the di-Arg motif. Furthermore, the ubiquitin R74A mutation shows a strong effect on activity, emphasizing the importance of its interaction with HOIP and of the resulting placement of the ubiquitin C terminus linked to the E2. The ubiquitin R72A mutant failed to form an UbCH5B~ubiquitin conjugate, thus preventing analysis. Coomassie-stained bands are in red. **f**, Overlap of the UbCH5B binding site on HOIP RING2L with the binding site of the acceptor ubiquitin. Left, UbCH5B~ubiquitin_{act} (orange/cyan) interaction with RING2L (green) from the HOIP RBR/E2~ubiquitin complex. Right, RING2L interaction with two ubiquitin molecules arranged in linear fashion, mimicking the HOIP RING2L~ubiquitin_{donor} to ubiquitin_{acceptor} (Ub_{don}, Ub_{acc}) transfer complex (PDB: 4LJP (ref. 14)). Despite the fact that the placement of the donor ubiquitin in the 4LJP structure results from a crystal contact, this ubiquitin exhibits a position identical to that of the activated ubiquitin bound to UBR2 in the HOIP RBR/UbCH5B~Ub complex. It should be noted that the UBR2 interaction with $h_{1,2}$ is missing because the RING2L from the crystal neighbour presenting the donor ubiquitin pushes $h_{1,2}$ into a different conformation. Importantly, in the HOIP RBR/UbCH5B~ubiquitin complex (left) the E2 binds RING2L in a region that overlaps with the binding site for the acceptor ubiquitin (Ub_{acc}, dark blue) in the RING2L~ubiquitin_{donor} to ubiquitin_{acceptor} transfer complex (right). This highlights how the E2~ubiquitin conjugate and the acceptor ubiquitin (which is the substrate of the E3 reaction) cannot bind the RBR at the same time, thus making a HECT-like transfer a requirement in the E3 ligase mechanism of RBR proteins.



Extended Data Figure 7 | Catalytic centre of the E2~ubiquitin/HOIP RBR E3 transfer complex. **a**, Close-up view of the catalytic centre of the transfer complex shows conservation of the contact conduits. Top left, close-up view of the catalytic centre in the HOIP/UbcH5B~ubiquitin transfer complex. Contact conduits 1 and 2 are highlighted with grey backgrounds. HOIP catalytic cysteine C885 is depicted in sphere representation. K85 replacing the catalytic cysteine (C85) in UbcH5B and ubiquitin G76 are displayed in stick representation, featuring the UbcH5B~ubiquitin linkage. Top right, model of the conduits in a Parkin/UbcH5B~ubiquitin complex. The structure of Parkin RING2 (from auto-inhibited Parkin, PDB: 4I1H (ref. 8)) was overlaid on that of HOIP RING2 indicating equivalent contact conduits. Bottom left, analogous model for HHARI using RING2 from auto-inhibited HHARI (PDB: 4KBL (ref. 11)). Bottom right, model of the conduits in a HOIP/UbcH7~ubiquitin complex. The model was generated from PDB entry 4Q5E (ref. 55) with UbcH7 (with the free catalytic cysteine C86 displayed) overlaid on UbcH5B of the HOIP/UbcH5B~ubiquitin transfer complex. The structure of the HOIP/UbcH5B~ubiquitin transfer complex and the other models depicted indicate a conservation of the contact conduits. Mechanistically, the conduits allow for the RBR catalytic cysteine and the E2 catalytic cysteine~ubiquitin linkage to be in close proximity, which serves as main driving force of the transesterification reaction. A reaction driven mainly by proximity is also in agreement with the chemical nature of the catalytic cysteine, which has a pK_a of ~8 for the free amino acid. This allows the cysteine to naturally deprotonate, without an absolute need for HOIP H887 (refs 10, 14), before attack of the ubiquitin G76 carbonyl function. In addition, the thioester linkage is far more labile than for example an amide bond, thus further facilitating a proximity-mediated reaction⁵⁶. However, in light of the geometric arrangement

observed, additional subtle catalytic contributions of H887 in supporting the transition state of the reaction and/or re-protonation of the E2 catalytic cysteine are in principle possible. This prospect is particularly intriguing because UbcH7 exhibits a potential break in conduit 2 (between H887 and H119), yet provides its 'own' histidine (H119) to the catalytic centre. **b**, Thioester transfer assays for HOIP contact conduit mutants. Left, thioester transfer assays show that the D983A and E976A mutations strongly affect activity. This is consistent with D983 forming the di-Arg binding motif in UBR2 (see Extended Data Fig. 6) and E976 bridging the three proteins in the complex. In contrast, the H887A mutation does not have a marked effect, in agreement with published results^{10,14}. The Q974A mutation also does not have a strong effect, pointing to a weak auxiliary function of this residue in support of the critical D983 (Coomassie-stained bands in red). Right, the UbcH5B R90A mutation shows a moderate yet significant effect, in line with the structure, which suggests a more pronounced effect for HOIP E976A than for UbcH5B R90A. Surprisingly, mutation of the catalytic D117 in UbcH5B, which is essential for classic RING-supported catalysis, shows a positive effect on the HECT-like thioester transfer, further emphasizing a separate mechanism for RBR HECT-like catalysis (as outlined in Fig. 2 and Extended Data Fig. 4). The gain of function of the D117A mutation also points to a trade-off for this E2 residue to participate in the classic RING-supported versus RBR HECT-like E2/E3 mechanisms (mean activity \pm s.e.m. ($n=3$), two-tailed unpaired Student's t -test; $**P < 0.01$; representative gels shown in Supplementary Fig. 1) **c**, Polyubiquitination assays for HOIP contact conduit alanine mutants. These assays show similar activity profiles as the thioester transfer assays except for the H887A mutation, which is essential for amide bond formation in the second transfer reaction^{10,14}.

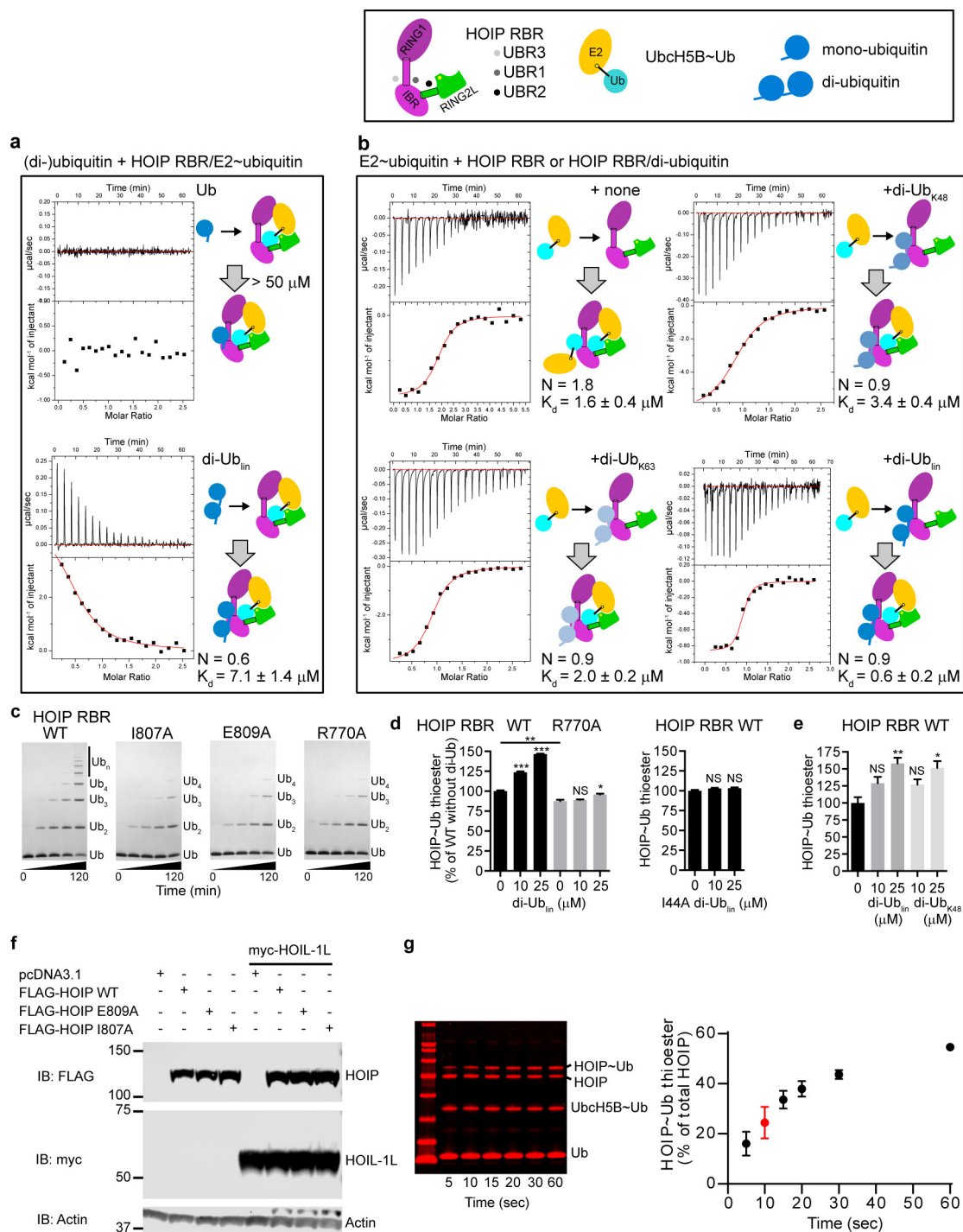


Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | The h_{E2} -IBR module contains an additional UBR (UBR3) that binds an allosteric ubiquitin using binding mode 2.

a, UBR3 binds the allosteric ubiquitin (Ub_{allo}) using binding mode 2. Cartoon depicting the overall features of binding mode 2 in UBR3. The binding of Ub_{allo} is largely analogous to the binding of the activated ubiquitin by UBR2 in the helix-IBR-fold of h_{L2} -RING2L (Extended Data Fig. 6a). Yellow squares indicate hydrophobic patches. **b**, Location of the UBR3/ Ub_{allo} interface in the overall complex and its relation to the UBR1/ Ub_{act} interface. The additional ubiquitin (Ub_{allo}) binds to UBR3 in the h_{E2} -IBR module immediately across UBR1 and the activated ubiquitin. The electron density map (feature-enhanced map⁵⁷, grey, contour level $\sigma = 1$) for the additional ubiquitin is also depicted on the right. **c**, Details of the UBR3/ Ub_{allo} interaction and its similarity to the interaction of phospho-ubiquitin (pUb) with *Pediculus humanus* Parkin (Ph-Parkin). Top left, close-up of HOIP UBR3. Ub_{allo} binds to the h_{E2} -IBR module with additional contacts to h_{E1} . Depicted are hydrophobic residues of HOIP (V789, L790, F796, W798 and I807) interacting with the ubiquitin canonical hydrophobic patch (L8, I44 and V70/not shown) and a second hydrophobic patch (L71 and L73/not shown), with the critical HOIP I807 emphasized. The di-Arg binding motif is also depicted, with E809 coordinating ubiquitin R72 and R74 and aligning the ubiquitin C terminus in a parallel manner with sheet $\beta 2$ of the IBR. A red arrow indicates that UBR3 sterically allows the binding of di-ubiquitin/polyubiquitin chains on the C-terminal side of the bound ubiquitin (see also Extended Data Fig. 9). Ubiquitin Ser65 is indicated for comparison with the pUb/Ph-Parkin structure (bottom). Top middle, close-up on the di-Arg binding motif. Top right, close-up on the additional contacts between h_{E1} and Ub_{allo} . HOIP R770 interacts with D766, which makes contacts to Y778, and the backbone carbonyl functions of ubiquitin K63 and E64. Bottom, the recent structure of phospho-ubiquitin bound to Ph-Parkin (PDB: 5CAW (ref. 13)) reveals a similar mechanism. Left, close-up with the residues corresponding to those in HOIP depicted. Middle, close-up of the di-Arg motif. Ph-Parkin D346 coordinates R72 similar to HOIP UBR3/ Ub_{allo} . The chemical tether introduced in the pUb/Ph-Parkin structure between the phospho-ubiquitin C terminus and a non-conserved Cys in Ph-Parkin (C349) shifts ubiquitin R74 away from Ph-Parkin D346 indicating that the di-Arg binding motifs of HOIP and Parkin undergo a similar interaction with ubiquitin or phospho-ubiquitin respectively. Right, the Ph-Parkin interaction equivalent to the HOIP h_{E1} interaction involves a ubiquitin phospho-serine 65 (pSer65) binding pocket in Ph-Parkin. Phospho-serine 65 is directly coordinated by R307 and Y314 and also H304, which is positioned similarly to HOIP D766. **d**, The binding of phospho-ubiquitin propagates the formation of UBR1 in Parkin. Overlay of Ub_{allo} (slate) bound to HOIP (magenta) and phospho-ubiquitin (light blue) bound to Ph-Parkin (grey blue). Ubiquitin binding propagates a straight conformation of h_{E2} and an opening of the IBR relative to the extended RING1 (only h_{E1} is shown). The tether introduced in the phospho-ubiquitin/Ph-Parkin interaction appears to exert a strain on IBR loop 2 and thus the IBR. This suggests that in the absence of the artificial

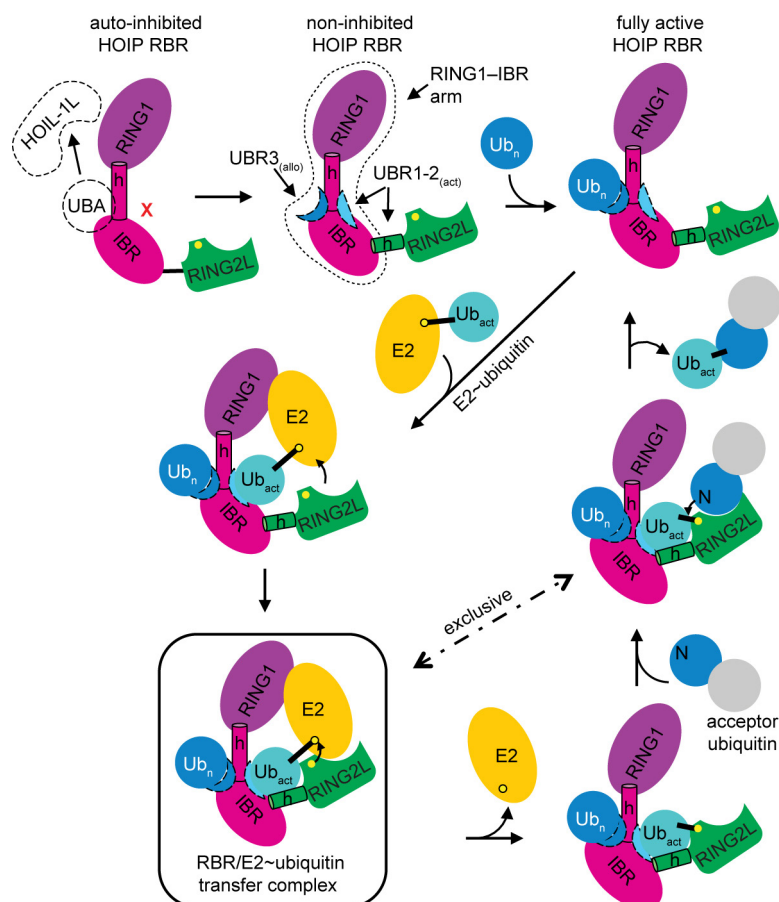
tether phospho-ubiquitin can propagate the formation of a fully functional UBR1 (binding Ub_{act}) in Parkin concomitantly to relieving the UBL autoinhibition, as elegantly demonstrated by Komander and colleagues¹³. **e**, The interaction of phospho-ubiquitin with a site analogous to UBR3 in Ph-Parkin causes a straight conformation of h_{E2} and reorientation of the IBR relative to RING1 as prerequisite to accommodate the activated ubiquitin. Left, full-length auto-inhibited human Parkin (PDB: 5C1Z (ref. 12)). Right, tethered phospho-ubiquitin in complex with Δ UBL-Ph-Parkin (PDB: 5CAW (ref. 13)). **f**, Conservation of UBR3 in HHARI. HHARI features a helix-IBR module similar to that of HOIP, with conserved hydrophobic patches and a polar residue (Q288) in the position of HOIP E809 that is in principle capable of binding the ubiquitin di-Arg motif. The h_{E2} -IBR region from auto-inhibited HHARI (PDB: 4KBL (ref. 11)) is displayed. As for auto-inhibited Parkin, a key difference from active HOIP is the kink in helix h_{E2} of auto-inhibited HHARI (see also panel **g** and Extended Data Fig. 5c). Taken together, these features indicate an overall similarity and the existence of a UBR3 that is allosterically linked to UBR1 in HOIP, Parkin and HHARI and potentially also other RBR proteins (see alignment in Supplementary Data 2). **g**, HHARI UBA domain (pink) binds HHARI RBR (light orange) in an equivalent position as Ub_{allo} binding to HOIP, but promotes an inhibitory conformation of RING1-IBR that cannot bind the activated ubiquitin (PDB: 4KBL (ref. 11)). **h**, UBR3 is a regulatory hotspot and UBR3/UBR1 crosstalk. Parkin auto-inhibition is facilitated by an UBL domain and is inherently linked to the equivalent of UBR3 and counteraction by phospho-ubiquitin binding. Different to Parkin, no structure of auto-inhibited HOIP RBR is available and thus the conformation of auto-inhibited HOIP is not known. However HHARI is, like HOIP, auto-inhibited by its UBA domain and the auto-inhibited structure of HHARI has been solved previously (PDB: 4KBL (ref. 11)). This structure reveals that unlike the UBL of Parkin, the UBA of HHARI directly utilizes the region of UBR3 for binding and auto-inhibition, which includes the kink in h_{E2} and a relative RING1-IBR positioning that is incompatible with binding of Ub_{act} as observed in this study, further underlying a regulatory 'hotspot' function of the RBR UBR3/UBR1. Left, the UBA domain (pink) of HHARI utilizes an anti-parallel β -sheet anchor with strand $\beta 2$ of the HHARI IBR (light orange) positioning the UBA to induce a kink in helix h_{E2} compared to its conformation in active HOIP (magenta) and counteracting the formation of a productive UBR3 and UBR1. Right, binding of the Ub_{allo} to UBR3 in mode 2 utilizes a parallel β -sheet anchor (centred around the di-Arg binding interaction) with strand $\beta 2$ of the IBR in active HOIP (magenta) inducing a straight conformation of h_{E2} and a conformation of UBR1 suited to bind the activated ubiquitin of the $E2 \sim Ub_{act}$ conjugate. Putative shifts for an analogous UBR1 formed in HHARI are indicated. Of note, the structure of auto-inhibited HOIP is not known and therefore the placement of the UBA in the schematic illustration of Extended Data Fig. 10 is deduced based on the similar auto-inhibition of HOIP and HHARI by their UBA domains, which still needs to be demonstrated.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | UBR3 interacts with di-ubiquitin and allosterically promotes E2~ubiquitin binding and HOIP RBR activation. **a**, ITC experiments analysing the binding of mono-ubiquitin or linear di-ubiquitin to an isolated HOIP RBR/E2~ubiquitin complex. While the binding of mono-ubiquitin is below the sensitivity of the experimental setting ($K_d > 50 \mu\text{M}$), the binding of linear di-ubiquitin exhibits a K_d of $7.1 \mu\text{M}$. **b**, ITC experiments analysing the binding of UbcH5B~ubiquitin to HOIP RBR in the absence and presence of different di-ubiquitin chains. Top left, the binding stoichiometry of UbcH5B~ubiquitin and wild-type HOIP RBR is $n = 1.8$, indicating that two UbcH5B~ubiquitin molecules interact with one RBR through UBR1/2 (catalytic binding) and UBR3 (binding of the ubiquitin moiety of the E2~ubiquitin conjugate) with a combined overall K_d of $1.6 \mu\text{M}$. The graph shows a single titration step, indicating 'crosstalk' between UBR3 and UBR1. Top right, the presence of K48-linked di-ubiquitin leads to 1:1 binding ($n = 0.9$) of UbcH5B~ubiquitin to HOIP RBR, indicating that di-ubiquitin occupies UBR3 and limits UbcH5B~ubiquitin binding to only its bona fide catalytic binding site (with ubiquitin binding to UBR1/2 and UbcH5B binding to RING1/RING2L). However, the presence of K48-linked di-ubiquitin results in the lowest affinity ($K_d = 3.4 \mu\text{M}$) for the binding of the conjugate to the RBR, indicating a negative effect of this linkage compared to the other di-ubiquitin entities tested. Bottom left, K63-linked di-ubiquitin has a more favourable effect on UbcH5B~ubiquitin binding ($K_d = 2.0 \mu\text{M}$). Bottom right, the strongest allosteric effect is observed in the presence of linear di-ubiquitin, which enables sub-micromolar binding ($K_d = 600 \text{ nM}$) of UbcH5B~ubiquitin. These results show that linear di-ubiquitin functions as a potent activator of HOIP RBR by binding to UBR3 (see also below and Fig. 4). While the structure depicts the interactions of one ubiquitin unit with UBR3, a second ubiquitin C-terminal to the UBR3-interacting ubiquitin may undergo further interactions with the IBR (as indicated by the arrow in Extended Data Fig. 8c). The cartoon representations summarize the configuration of each ITC experiment. **c**, Polyubiquitination assays of UBR3 mutants. The HOIP RBR I807A, E809A and R770A mutants exhibit a marked reduction in activity, supporting the importance of UBR3 in HOIP function. **d**, Activation of HOIP RBR by di-ubiquitin. While wild-type HOIP RBR is activated by the presence of increasing concentrations of wild-type linear di-ubiquitin, wild-type linear di-ubiquitin only has a

weak effect on activation of the UBR3 R770A mutant (similar to the I807A and E809A mutants in Fig. 4b). Additionally the di-ubiquitin mutant I44A, which is mutated at a critical UBR3-interacting residue in both ubiquitin units, does not have an activating effect on wild-type HOIP RBR thioester activity (mean activity \pm s.e.m. ($n = 3$), one-way ANOVA followed by Tukey's post hoc test; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; NS, not significant; representative gels shown in Supplementary Fig. 1). **e**, Effect of linear versus K48-linked di-ubiquitin on HOIP RBR thioester transfer activity. In contrast to the ITC binding studies in **b**, linear and K48-linked di-ubiquitin are both able to increase the thioester transfer activity of HOIP RBR (although the experimental setup necessary to investigate the K48-linkage resulted in larger error; see Methods; mean activity \pm s.e.m. ($n = 3$), one-way ANOVA followed by Tukey's post hoc test; $*P < 0.05$; $**P < 0.01$; NS, not significant; representative gels shown in Supplementary Fig. 1). These results show an UBR3-dependent activating effect of di-ubiquitin, and thus potentially of polyubiquitin chains. However, whether HOIP UBR3 acts as a universal ubiquitin sensor or has a preference for linear ubiquitin over other types of linkage needs to be further examined through careful investigations also including full-length proteins of the LUBAC in cellular settings. Additionally, although there is a substantial gap between UBR3 and the position of the acceptor ubiquitin, longer acceptor ubiquitin chains might be able to bridge this gap and mediate a cooperative effect between the two sites. This would be consistent with a recent publication showing that the presence of K63-linked ubiquitin chains is frequently necessary for the formation of linear polyubiquitin chains⁵⁸. **f**, Protein expression levels for the NF- κ B reporter assays in cells shown in Fig. 4d. Shown are anti-Flag immunoblots of wild-type HOIP and mutants and anti-myc immunoblots of HOIL-1L, demonstrating similar protein expression levels in different cell lysates. Lysates were also probed by immunoblotting for actin as a loading control. Uncropped blots are shown in Supplementary Fig. 1. **g**, Time course of HOIP~ubiquitin thioester transfer assay. Left, SDS-PAGE showing time course of HOIP~ubiquitin thioester transfer assay. Coomassie-stained bands in red visualized using LI-COR Odyssey at 700 nm. Right, plot of quantified HOIP~ubiquitin thioester transfer assay time-course (mean \pm s.e.m., $n = 2$). The 10-s time point used in the end-point assays throughout the study is highlighted in red.



Extended Data Figure 10 | Schematic of RBR mechanism: HOIP RBR activation and E3 ligase cycle. HOIP RBR is initially auto-inhibited by its UBA domain. Sequestration of the auto-inhibitory HOIP UBA domain by HOIL-1L^{5-7,18} releases the conformational restraint exerted by the UBA, allowing formation of UBR1 and UBR3. Binding of a ubiquitin entity such as a linear ubiquitin chain to UBR3 stabilizes the active conformation of UBR1 and the RING1-IBR arm, facilitating binding of the E2~ubiquitin conjugate. In the subsequent HOIP/E2~ubiquitin transfer complex, the E2~ubiquitin conjugate is engaged in a clamp-like manner bringing the RBR active cysteine and the E2~ubiquitin thioester in close proximity,

ultimately leading to the transfer of the ubiquitin to the RBR cysteine. The E2 then vacates the complex, freeing the site for binding of the acceptor ubiquitin, the N-terminal amine of which attacks the RBR thioester^{7,14}. Once the ubiquitin chain linkage is formed, the ubiquitinated substrate/growing ubiquitin chain must exit RING2L to enable binding of a new E2~ubiquitin conjugate for the next loading of the RBR in the HECT-like E3 ligase cycle. The growing ubiquitin chain could be retained near the RBR by the HOIP NZF domains, HOIL-1L or SHARPIN^{15-17,59}, directly linking the HECT-like mechanism to co-operative processes within the LUBAC.

Structure of transcribing mammalian RNA polymerase II

Carrie Bernecky¹, Franz Herzog², Wolfgang Baumeister³, Jürgen M. Plitzko³ & Patrick Cramer¹

RNA polymerase (Pol) II produces messenger RNA during transcription of protein-coding genes in all eukaryotic cells. The Pol II structure is known at high resolution from X-ray crystallography for two yeast species^{1–3}. Structural studies of mammalian Pol II, however, remain limited to low-resolution electron microscopy analysis of human Pol II and its complexes with various proteins^{4–10}. Here we report the 3.4 Å resolution cryo-electron microscopy structure of mammalian Pol II in the form of a transcribing complex comprising DNA template and RNA transcript. We use bovine Pol II, which is identical to the human enzyme except for seven amino-acid residues. The obtained atomic model closely resembles its yeast counterpart, but also reveals unknown features. Binding of nucleic acids to the polymerase involves ‘induced fit’ of the mobile Pol II clamp and active centre region. DNA downstream of the transcription bubble contacts a conserved ‘TPSA motif’ in the jaw domain of the Pol II subunit RPB5, an interaction that is apparently already established during transcription initiation⁷. Upstream DNA emanates from the active centre cleft at an angle of approximately 105° with respect to downstream DNA. This position of upstream DNA allows for binding of the general transcription elongation factor DSIF (SPT4–SPT5) that we localize over the active centre cleft in a conserved position on the clamp domain of Pol II. Our results define the structure of mammalian Pol II in its functional state, indicate that previous crystallographic analysis of yeast Pol II is relevant for understanding gene transcription in all eukaryotes, and provide a starting point for a mechanistic analysis of human transcription.

To determine the high-resolution structure of mammalian Pol II, we prepared the bovine enzyme from calf thymus (Methods). Purified bovine Pol II contained all 12 subunits (RPB1–RPB12) in apparently stoichiometric amounts and bound to human Gdown1, an additional metazoan-specific Pol II subunit¹¹ (Extended Data Fig. 1a). The polymerase also bound a synthetic DNA–RNA scaffold to form an elongation complex (EC) that was active in extending the RNA transcript (Extended Data Fig. 1b, c). Crystallization trials were unsuccessful, but we could determine a high-resolution structure of the bovine Pol II EC by cryo-electron microscopy (cryo-EM) and single particle reconstruction (Methods).

The EC sample was used for cryo-EM data collection with a K2 direct electron detection device (Gatan). These data revealed particles of the expected size that resulted in defined 2D class averages (Extended Data Fig. 1d, e). Unsupervised 3D classification of 409,401 particle images led to a reconstruction of the Pol II EC from 264,134 particles at an overall resolution of 3.4 Å and a local resolution of 3.0 Å in the active centre (EC1, Extended Data Figs 2 and 3). Upstream DNA and the RPB4–RPB7 stalk were flexible, but alternative sorting of particles led to EC reconstructions with improved density for the RPB4–RPB7 stalk (EC2; 219,265 particles) and for upstream DNA (EC3; 184,122 particles) at resolutions of 3.6 Å and 3.7 Å, respectively.

Density for Gdown1 was not observed in these reconstructions, probably because Gdown1 is flexibly tethered to Pol II and requires interactions with additional factors to adopt a defined location that could be detected by cryo-EM.

The vast majority of imaged particles encompassed the 12-subunit Pol II EC. A minor fraction of EC particles with a similar angular distribution displayed less density for the jaw-lobe region, indicating mobility as suggested by a previous low-resolution EM study of human Pol II⁴. An additional fraction of particles lacking nucleic acids was also observed (Extended Data Fig. 2). These particles contained free Pol II with a flexible clamp, also consistent with previous observations⁴. The corresponding reconstruction showed that parts of the active centre region and hybrid-binding domain were also flexible. This flexibility was only observed in the absence of nucleic acids, indicating that Pol II undergoes an induced fit when binding to nucleic acids. In the absence of nucleic acids, fork loop 2 was rearranged, covering a positively charged patch that interacts with the downstream edge of the single-stranded region of the DNA non-template strand in the EC.

The mobility of the clamp was previously inferred from crystallographic studies of yeast Pol II. Crystal structures of the ten-subunit core Pol II, lacking the Rpb4–Rpb7 stalk, trapped the clamp in two different open positions², whereas binding of nucleic acids resulted in a defined, closed position of the clamp¹² that corresponded to the one observed here. The same closed position was observed in the absence of nucleic acids in crystals of the complete Pol II containing the stalk^{13,14}, although we observed here mobility of the clamp even in the presence of the stalk. In previous crystals, the stalk was involved in crystal contacts, but its conformation nevertheless matched the one seen in EM studies of yeast Pol II¹⁵. In the mammalian Pol II EC presented here, the stalk adopts a different orientation, which resembles that seen in a yeast EM structure of the core initiation complex bound by the core Mediator¹⁵. Taken together, crystallization traps Pol II conformations that exist in solution, whereas EM has the potential to unveil multiple conformations, providing insights into the conformational dynamics of the complex.

The cryo-EM density for EC1 revealed protein side chains and single nucleotides, and was comparable in quality to crystallographic maps calculated with refined model phases at a similar resolution (Fig. 1a). We built an atomic model based on the homologous yeast Pol II structure and refined the model in real space. The resulting structure included 95% of all Pol II residues (omitting the flexible carboxy (C)-terminal domain of Pol II subunit RPB1) and showed very good stereochemistry (Extended Data Table 1). The structure represents one of only four asymmetric macromolecular structures with a molecular mass of around 1 MDa or less that were thus far resolved by cryo-EM at near-atomic resolution (<4 Å)^{16–19}.

The overall structure of the mammalian Pol II EC (Fig. 1b) resembles the yeast EC crystal structures^{12,20}, but also shows differences

¹Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Am Faßberg 11, 37077 Göttingen, Germany. ²Gene Center Munich, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 Munich, Germany. ³Max Planck Institute for Biochemistry, Department of Molecular Structural Biology, Am Klopferspitz 18, 82152 Martinsried, Germany.

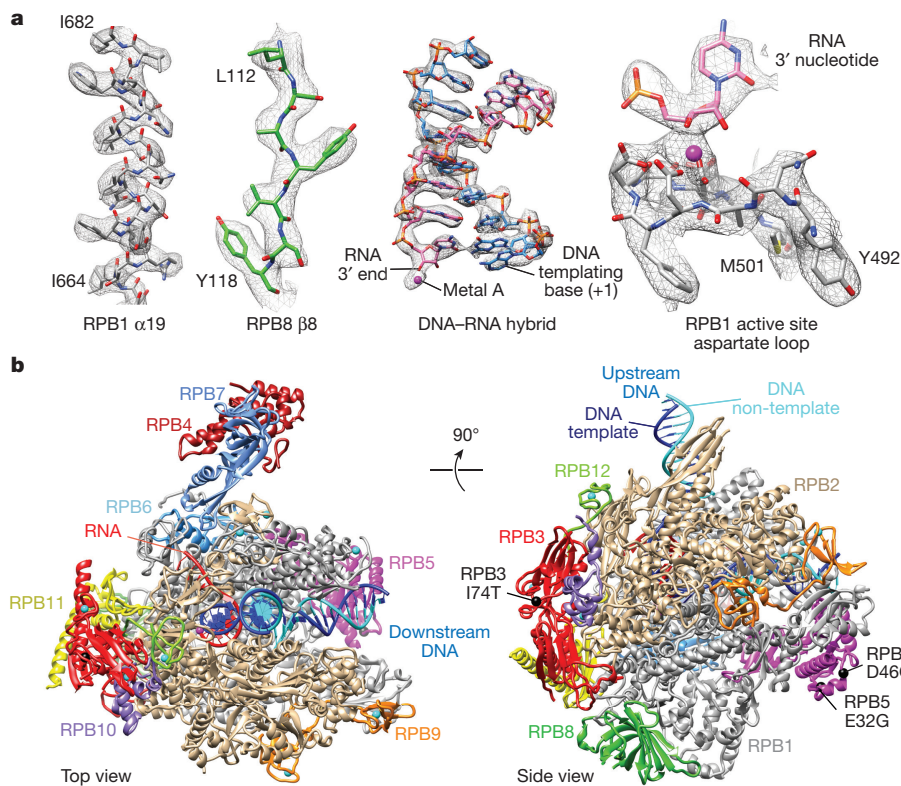


Figure 1 | Cryo-EM structure of mammalian Pol II EC at 3.4 Å resolution. **a**, Representative regions of the cryo-EM density for EC1 with the refined model superimposed. Depicted are (from left to right) RPB1 helix α 19, RPB8 strand β 8, the DNA-RNA hybrid, and the active site aspartate loop with the bound catalytic metal ion A and the 3'-nucleotide of the RNA transcript. **b**, Ribbon model. The views correspond to the previously used 'top' and 'side' views of yeast Pol II and are related by a 90° rotation around a horizontal axis. Black spheres indicate the location of residues that are not identical between bovine and human Pol II, bovine indicated second (three out of seven residues, the remaining four are disordered). The final model lacked several short surface loops and flexible N-terminal residues.

in surface regions of the polymerase (Extended Data Fig. 4a, b and Extended Data Table 2). First, mammalian Pol II exhibits an insertion in subunit RPB8 that contacts the RPB1 foot domain, which is rearranged compared with the yeast enzyme (Fig. 2a). The *Saccharomyces cerevisiae* foot domain forms a transient contact with the coactivator complex Mediator¹⁵. It is possible that these differences affect mammalian Mediator binding. Second, mammalian Pol II contains two insertions in the region of the Pol II pore that accommodates the RNA cleavage factor TFIIS beneath the active site (Fig. 2b). The insertion β 20– β 21 contains a new negatively charged helix (called here α 16–2) that may interact with a positively charged region of TFIIS domain 3. The insertion in strand β 32 is in a position to interact with the linker between TFIIS domains 2 and 3. This may be relevant for binding factors with TFIIS-like domains that are present in mammalian cells, but not in yeast. Third, the structure reveals that the RPB5 jaw domain is rotated by $\sim 5^\circ$ towards downstream DNA (Fig. 2c, d and Extended Data Fig. 4c). Fourth, the EC1 map reveals nearly continuous density for the trigger loop in the active centre. The trigger loop adopts an open conformation, similar to the trapped state observed in the presence of backtracked RNA²¹ and an open state observed in a recent X-ray structure of a yeast Pol II EC²² (Extended Data Fig. 4d).

With respect to the nucleic acids, downstream DNA enters the active centre cleft and unwinds before the active site when the template strand passes over the bridge helix. The RNA transcript emanates from the active site aspartate loop and forms a hybrid with the DNA template strand. RNA separates at the upstream end of the hybrid and passes beneath the lid loop through the proposed RNA exit tunnel to reach the Pol II surface near the dock domain when it is 14 nucleotides long. Density for exiting RNA extends beyond this point, but could not be modelled owing to its increasing flexibility on the enzyme surface. The upstream DNA duplex also shows weaker density, consistent with its known mobility²⁰, but adopts a defined position in the EC3 reconstruction (Fig. 3a).

Most interactions between Pol II and the nucleic acids occur in the region of the DNA-RNA hybrid and generally involve residues that are either identical or conserved between yeast and human enzymes (Fig. 3b, c). Density for an insertion within the RPB2 lobe (loop β 9– β 10) approached the downstream DNA. This loop is conserved in metazoa

but not yeast, and contains two lysine residues near DNA. Further downstream, DNA contacts the Pol II jaw domain of subunit RPB5, which was moved by up to 3 Å compared with the yeast EC. The contact is made by a 'TPSA' motif comprising four conserved amino-acid residues that form the first turn of helix α 6 in RPB5 (Fig. 2d). A contact of downstream DNA with RPB5 is already established in the closed

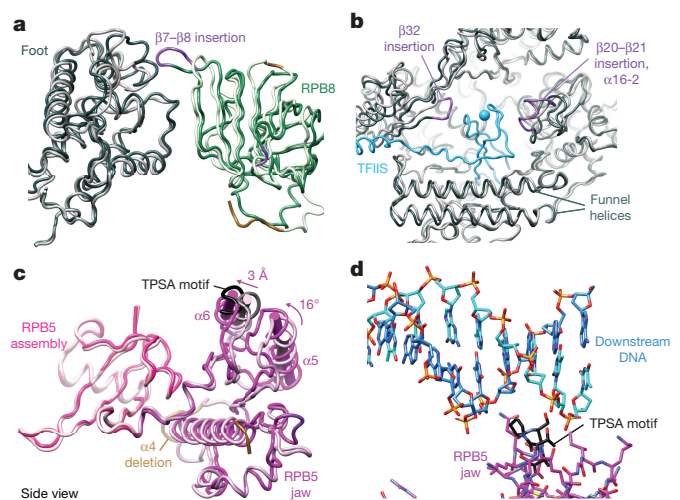
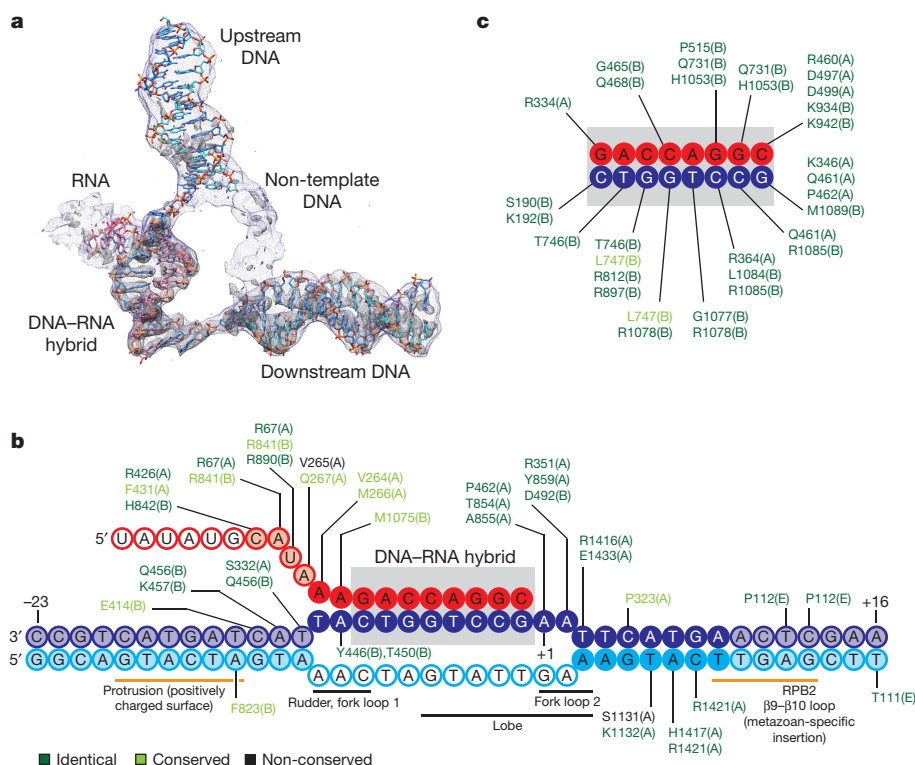


Figure 2 | Selected mammalian-specific features that differ from yeast Pol II (PDB accession number 1Y1W). **a**, An insertion in RPB8 and a rearrangement in the RPB1 foot domain in bovine Pol II lead to increased contacts in this region. Bovine Pol II is shown in darker colours (RPB1 dark grey; RPB8 dark green) and yeast Pol II is shown in lighter colours (RPB1 light grey; RPB8 light green). Insertions within bovine Pol II are coloured purple and insertions within yeast Pol II are coloured brown. **b**, Two insertions in RPB1 near the pore and TFIIS-binding region. Colouring is as in **a**, and TFIIS is in blue. **c**, Changes in the RPB5 jaw domain compared with the structurally conserved RPB5 assembly domain. Insertions are coloured as in **a**. **d**, Contact of the TPSA motif (black) in RPB5 with downstream DNA (blue).

**Figure 3 | Pol II-nucleic-acid interactions.**

a, EM density for nucleic acids. The density for the B-factor sharpened Pol II EC1 map is shown in grey mesh. The density for the unsharpened Pol II EC3 map is shown in blue mesh. The view corresponds to the previously used 'side' view. **b**, Pol II residues interacting with nucleic acids. Schematic of the nucleic-acid scaffold with RNA in red, template DNA in blue, and non-template DNA in cyan. Dark filled circles represent nucleotides that were well resolved both in this study and in the previous yeast EC X-ray structure (PDB accession number 1Y1W). Lighter filled circles represent nucleotides that could be additionally modelled. Unfilled circles represent nucleotides that were not modelled. Residues within 4 Å of nucleic acids are shown, with the subunit chain identifier (A–L for RPB1–RPB12) in parenthesis. Residues that are identical between yeast and bovine Pol II are in green, similar residues are yellow-green, and non-conserved residues are black. The polymerase domains in proximity to single-stranded non-template DNA are in black, whereas domains in proximity to upstream or downstream DNA are denoted in orange. **c**, Pol II residues interacting with the DNA-RNA hybrid.

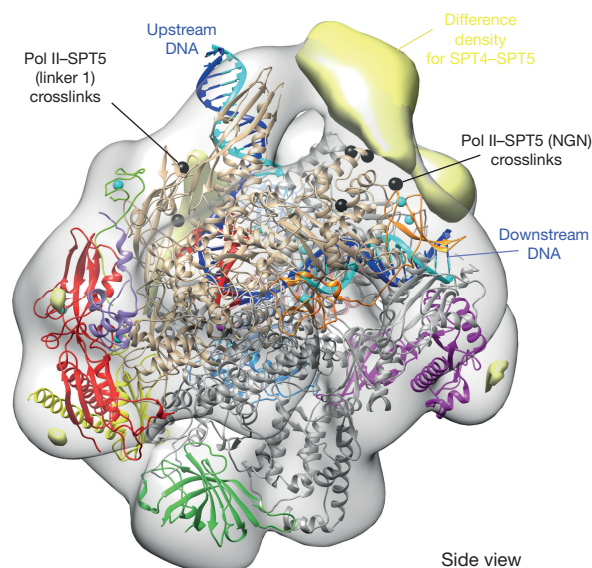
promoter complex and an open complex mimic⁷, indicating that it is maintained during the transition from initiation to elongation (Extended Data Fig. 4e). These observations are consistent with the model that downstream DNA is moved into the Pol II cleft by a translocation along RPB5 upon TFIIH action and DNA opening^{7,23}. Studies of ten-subunit yeast Pol II have revealed a similar contact with downstream DNA^{22,24}.

The orientation of upstream DNA with Pol II deviates from previous observations made in the yeast system. Upstream DNA emanates from the Pol II cleft between the clamp and protrusion domains at an angle of $\sim 105^\circ$ with respect to downstream DNA, compared with the previously described $\sim 90^\circ$ on the basis of fluorescence resonance energy transfer measurements²⁵ and $\sim 130^\circ$ on the basis of a recent X-ray structure of a 10-subunit yeast Pol II with a larger, 15-nucleotide DNA mismatch bubble²². The deviation in the angle of the upstream DNA could be a result of the larger mismatch region, the redistribution of positive charges on the protrusion between yeast and mammals, or an altered conformation of loop $\beta 28$ – $\beta 29$ on the RPB2 wall domain that approaches the minor groove of upstream DNA.

The observed position of upstream DNA is compatible with binding of the elongation factor DSIF (SPT4–SPT5) to the clamp as modelled on the basis of the archaeal and yeast complexes, whereas the previous position of upstream DNA clashed with DSIF^{26–28}. To test our model, we prepared an EC with bound DSIF and determined the location of DSIF with the use of negative stain EM and crosslinking coupled to mass spectrometry. As predicted, SPT4 and the NusG amino (N)-terminal (NGN) domain of SPT5 occupied the top of the cleft, bridging between the clamp on one side and the lobe and protrusion on the other side of Pol II (Fig. 4 and Supplementary Table 1). Modelling on the basis of the previous X-ray structure of an archaeal clamp complex with SPT4–SPT5 places positively charged regions of the SPT5 NGN domain in close proximity to the observed position of upstream DNA and non-template DNA (Extended Data Fig. 5), consistent with a role of this conserved family of elongation factors in stabilizing the upstream edge of the transcription bubble^{29,30}.

In summary, high-resolution structural analysis of a mammalian Pol II has become possible owing to advances in cryo-EM with the use of crystallization-grade preparations of the enzyme that have been

available for years but never produced diffraction-quality crystals. The cryo-EM density map for the mammalian Pol II EC resembles in quality and resolution the best crystallographic maps obtained for the corresponding yeast complex. Comparison of the mammalian cryo-EM structure with the yeast X-ray structures reveals differences between these highly similar enzymes and provides insights into the conformational control of enzyme function. Most importantly, this work provides the basis for future structure determination of mammalian Pol II transcription complexes with mammalian-specific factors.

**Figure 4 | Location of DSIF on the Pol II EC.** Negative stain EM and crosslinking of the Pol II-DSIF EC. A difference map between the negative stain Pol II-DSIF EC reconstruction at 26 Å resolution and the EC model is shown in yellow. Pol II residues observed to crosslink to DSIF are shown as black spheres. Although present in the Pol II-DSIF EC, nucleic acids are not visible owing to the use of negative stain.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 July; accepted 24 November 2015.

Published online 20 January 2016.

- Cramer, P. *et al.* Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649 (2000).
- Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).
- Spähr, H., Calero, G., Bushnell, D. A. & Kornberg, R. D. *Schizosaccharomyces pombe* RNA polymerase II at 3.6-Å resolution. *Proc. Natl Acad. Sci. USA* **106**, 9185–9190 (2009).
- Kostek, S. A. *et al.* Molecular architecture and conformational flexibility of human RNA polymerase II. *Structure* **14**, 1691–1700 (2006).
- Kassube, S. A., Jinek, M., Fang, J., Tsutakawa, S. & Nogales, E. Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. *Nature Struct. Mol. Biol.* **20**, 892–899 (2013).
- Kassube, S. A. *et al.* Structural insights into transcriptional repression by noncoding RNAs that bind to human Pol II. *J. Mol. Biol.* **425**, 3639–3648 (2013).
- He, Y., Fang, J., Taatjes, D. J. & Nogales, E. Structural visualization of key steps in human transcription initiation. *Nature* **495**, 481–486 (2013).
- Bernecky, C., Grob, P., Ebmeier, C. C., Nogales, E. & Taatjes, D. J. Molecular architecture of the human Mediator–RNA polymerase II–TFIIF assembly. *PLoS Biol.* **9**, e1000603 (2011).
- Bernecky, C. & Taatjes, D. J. Activator–mediator binding stabilizes RNA polymerase II orientation within the human mediator–RNA polymerase II–TFIIF assembly. *J. Mol. Biol.* **417**, 387–394 (2012).
- Wu, Y. M. *et al.* Regulation of mammalian transcription by Gdown1 through a novel steric crosstalk revealed by cryo-EM. *EMBO J.* **31**, 3575–3587 (2012).
- Hu, X. *et al.* A Mediator-responsive form of metazoan RNA polymerase II. *Proc. Natl Acad. Sci. USA* **103**, 9506–9511 (2006).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Armache, K. J., Kettenberger, H. & Cramer, P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl Acad. Sci. USA* **100**, 6964–6968 (2003).
- Bushnell, D. A. & Kornberg, R. D. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl Acad. Sci. USA* **100**, 6969–6973 (2003).
- Plaschka, C. *et al.* Architecture of the RNA polymerase II–Mediator core initiation complex. *Nature* **518**, 376–380 (2015).
- Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Atomic structure of the APC/C and its mechanism of protein ubiquitination. *Nature* **522**, 450–454 (2015).
- Liang, B. *et al.* Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell* **162**, 314–327 (2015).
- Urnavicius, L. *et al.* The structure of the dynein complex and its interaction with dynein. *Science* **347**, 1441–1446 (2015).
- Bai, X. C. *et al.* An atomic structure of human γ -secretase. *Nature* **525**, 212–217 (2015).
- Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIS. *Mol. Cell* **16**, 955–965 (2004).
- Martinez-Rucobo, F. W. & Cramer, P. Structural basis of transcription elongation. *Biochim. Biophys. Acta* **1829**, 9–19 (2013).
- Barnes, C. O. *et al.* Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol. Cell* **59**, 258–269 (2015).
- Fishburn, J., Tomko, E., Galburt, E. & Hahn, S. Double-stranded DNA translocase activity of transcription factor TFIIF and the mechanism of RNA polymerase II open complex formation. *Proc. Natl Acad. Sci. USA* **112**, 3961–3966 (2015).
- Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954 (2006).
- Andrecka, J. *et al.* Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res.* **37**, 5803–5809 (2009).
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
- Li, W., Giles, C. & Li, S. Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res.* **42**, 7069–7083 (2014).
- Hirtreiter, A. *et al.* Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38**, 4040–4051 (2010).
- Belogurov, G. A., Sevostyanova, A., Svetlov, V. & Artsimovitch, I. Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol. Microbiol.* **76**, 286–301 (2010).
- Sevostyanova, A. & Artsimovitch, I. Functional analysis of *Thermus thermophilus* transcription factor NusG. *Nucleic Acids Res.* **38**, 7432–7445 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Wolf for providing bovine thymus. We thank D. Pöhlmann for preparation of human DSIF, and M. Hemann for generation of peptides for mass spectrometry. We thank J. Schuller and C. Plaschka for assistance with EM data collection. We thank D. Tegunov and C. Plaschka for discussions about EM data processing. We thank D. Haselbach and W. Liu for assistance with cryo-EM grid preparation. We thank O. Berninghausen for providing access to a Tecnai Spirit microscope. C.B. was supported by the Alexander von Humboldt Foundation. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860), the Advanced Grant ‘TRANSIT’ of the European Research Council, and the Volkswagen Foundation.

Author Contributions C.B. designed and performed all experiments and data analysis. F.H. supervised mass spectrometry. W.B. and J.P. provided access to a high-end EM facility. J.P. supervised EM data collection. P.C. designed and supervised research. C.B. and P.C. prepared the manuscript.

Author Information Cryo-electron microscopy densities have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-3218 (EC1), EMD-3219 (EC2), and EMD-3220 (EC3). Model coordinates have been deposited in the Protein Data Bank (PDB) under accession number 5FLM. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (pcramer@mpibpc.mpg.de).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Purification of bovine Pol II. Bovine Pol II was prepared as described¹¹ with modifications. Unless otherwise noted, all steps were completed at 4°C. Protease inhibitors included 1 mM PMSF, 1 mM benzamidine, 60 µM leupeptin, and 200 µM pepstatin. Calf thymus was homogenized for 3 min in buffer A (50 mM Tris, pH 7.9 at 4°C, 1 mM EDTA, 10 µM ZnCl₂, 10% glycerol, 1 mM DTT, protease inhibitors) using a 21 blender (Waring). The homogenized material was centrifuged and the supernatant filtered through two layers of Miracloth. A 5% solution of polyethyleneimine, pH 7.9 at 25°C, was added to a final concentration of 0.02%, and the material was stirred for 10 min then centrifuged. The resulting pellets were washed with buffer A before resuspension in buffer A (0.15 M ammonium sulfate). After centrifugation, the conductivity of the supernatant was adjusted to that of buffer A (0.2 M ammonium sulfate), and the resulting material was loaded on a 225-ml MacroPrepQ column equilibrated in buffer A (0.2 M ammonium sulfate). The column was washed with two column volumes of buffer A (0.2 M ammonium sulfate), followed by Pol II elution with buffer A (0.4 M ammonium sulfate). The eluate was precipitated by addition of finely ground ammonium sulfate added to 50% saturation, and pellets were collected by centrifugation. The pellets were resuspended in buffer A, and the conductivity was adjusted to that of buffer A (0.15 M ammonium sulfate). The material was clarified by centrifugation, and further purified using a 5-ml gravity flow column of 8WG16 (αRPB1 CTD) antibody-coupled sepharose equilibrated in buffer A (0.15 M ammonium sulfate). After application of the input material, the antibody column was washed with five column volumes of buffer A (0.5 M ammonium sulfate), sealed, and allowed to equilibrate to room temperature (20–25°C) for 15 min. Pol II was eluted using buffer A (0.5 M ammonium sulfate, 50% (v/v) glycerol), and Pol-II-containing fractions were immediately mixed with buffer A (2 mM DTT, lacking glycerol and protease inhibitors). The diluted material was centrifuged and subjected to anion exchange chromatography using a UNO-Q column equilibrated in buffer A (0.1 M ammonium sulfate, 2 mM DTT, lacking protease inhibitors). Pol II was eluted using a linear gradient from 0.1 M to 0.5 M ammonium sulfate in buffer A (2 mM DTT, lacking protease inhibitors). For the purification of 12-subunit bovine Pol II, the Gdown1-free Pol II fraction was applied to a Sephacryl S-300 HiLoad sizing column equilibrated in buffer B (150 mM NaCl, 5 mM HEPES pH 7.25 at 25°C, 10 µM ZnCl₂, 10 mM DTT). For the purification of bovine Pol II containing Gdown1, the Gdown1-free Pol II fraction was incubated with a 3× molar excess of human Gdown1 for 1 h at 4°C before application to the Sephacryl S-300 HiLoad sizing column. Pol-II-containing fractions were concentrated using a 100-kDa cutoff Amicon concentrator to a final concentration of 2–4 mg ml⁻¹.

Preparation of recombinant human proteins. Gene-optimized human Gdown1 (Life Technologies) was cloned into pOPINB (N-terminal His₆ tag and 3C protease site). After transformation, *Escherichia coli* BL21(DE3)RIL cells were grown at 37°C in Lysogeny broth (LB) medium to an absorbance at 600 nm, $A_{600\text{nm}}$, of 0.5 before protein expression with 0.5 mM IPTG for 3–4 h at 37°C. Subsequent steps were completed at 4°C unless otherwise noted. Cells were lysed by sonication in buffer C (50 mM HEPES pH 7.5 (25°C), 300 mM NaCl, 1 mM CaCl₂, 10% glycerol) supplemented with 10 mM imidazole, 1 mM PMSF, 1 mM benzamidine, 1 mM sodium metabisulfite, 1 mM DTT, and 2 µg ml⁻¹ DNase I. Cleared lysate was subjected to affinity chromatography using Ni-NTA agarose (Qiagen), and excess chaperone was removed by washing the resin with a 5 mM ATP and 2 mg ml⁻¹ denatured *E. coli* protein wash at room temperature in buffer C supplemented as above containing 30 mM imidazole. Protein was eluted with buffer C supplemented as above, but lacking DNase I and containing 250 mM imidazole. Elutions were exchanged into buffer C supplemented with 10 mM imidazole and 1 mM DTT via a PD10 desalting column, followed by 3C protease cleavage at 4°C overnight. Cleaved Gdown1 was subjected to reverse chromatography (Ni-NTA agarose) followed by dilution with buffer D (50 mM HEPES pH 7.5 (25°C), 1 mM CaCl₂, 10% glycerol, 2 mM DTT) to a conductivity of buffer D containing 0.05 M NaCl. Diluted protein was subjected to cation exchange chromatography (MonoS 5/50) to remove additional chaperone, and eluted with a linear gradient from 0.05 M to 0.5 M NaCl in buffer D. The conductivity of the Gdown1-containing fractions was again adjusted to that of buffer D containing 0.05 M NaCl, and applied to a MonoQ 5/50 anion exchange column. Gdown1 was eluted using a linear gradient from 0.05 M to 0.5 M NaCl in buffer D. Fractions containing purified Gdown1 were pooled, resulting in a final concentration of 1–1.5 mg ml⁻¹. Yield was approximately 2.5 mg per 21 of *E. coli* culture.

Purification of human SPT4 and SPT5 was as described³¹, with adaptations. Gene-optimized human SPT5 (pMK vector, no tag) and SPT4 were purchased from Life Technologies, and SPT4 was recloned into pOPINJ (N-terminal HIS₆ and GST tags followed by a 3C protease cleavage site). SPT4 and SPT5 vectors

were co-transformed into *E. coli* BL21(DE3)RIL cells, which were then grown at 37°C in LB medium supplemented with 10 µM ZnCl₂ to $A_{600\text{nm}} = 0.6$. Expression was induced with 1 mM IPTG for 18 h at 18°C. Cells were lysed by sonication in buffer E (25 mM Tris pH 7.4 (4°C), 500 mM NaCl, 10 µM ZnCl₂, 5 mM DTT) supplemented with 5 mM imidazole and protease inhibitors (1 mM PMSF, 1 mM benzamidine, 60 µM leupeptin, and 200 µM pepstatin). Soluble material was passed over a Ni-NTA agarose column and washed with ten column volumes each of buffer E supplemented with 20 mM or 40 mM imidazole before elution in buffer E supplemented with 300 mM imidazole. Eluted protein was cleaved with 3C protease during overnight dialysis (4°C) against buffer E, then subjected to reverse chromatography. Protein was passed over a HiTrap Q HP anion exchange column to remove DNA, and the flow-through fraction containing SPT4/5 was concentrated using a 50-kDa cutoff Amicon concentrator to 1–4 mg ml⁻¹.

Elongation complex preparation. The nucleic-acid scaffold (Metabion) was only slightly modified from what was used for a yeast Pol II EC crystal structure²⁰ by mutation of the DNA templating base from C to A to generate a fully mismatched open bubble and by removing the upstream and downstream CC overhangs. A 1.5× molar excess of pre-annealed template DNA (sequence 5'-AAGCTCAAGTACTTAAAGCCTGGTCATTACTAGTACTGCC-3'), non-template DNA (sequence 5'-GGCAGTACTAGTAACTAGTATTGAAAGTACTTGAGCTT-3'), and RNA (sequence 5'-UAUAUGCAUAAAGACCAGGC-3') were incubated with Pol II-Gdown1 at 4°C for 10 min, then 20°C for 15 min. To increase the randomness of Pol II EC particle orientations, the resulting complex (0.85 µM) was crosslinked with 3 mM BS3 (Thermo) for 30 min at 30°C, then quenched with 50 mM ammonium bicarbonate. Crosslinked complex was applied to a Superdex 200 increase 10/300 GL column equilibrated in buffer B (150 mM NaCl, 5 mM HEPES pH 7.25 at 25°C, 10 µM ZnCl₂, 10 mM DTT). The nucleic-acid-containing peak was concentrated to ~0.3 mg ml⁻¹ as described above and used immediately for cryo-EM grid preparation.

Electron microscopy. Four microlitres of sample were applied to glow-discharged Quantifoil R 3.5/1 holey carbon grids, which were then blotted and plunge-frozen in liquid ethane using a Vitrobot (FEI). Data were acquired using an FEI Titan Krios operated in energy-filtered transmission electron microscopy (EFTM) mode at 300 kV equipped with a Gatan K2 Summit direct detector. Automated data collection was performed using the TOM toolbox³². Movie images were collected at a nominal magnification of ×37,000 (1.35 Å per pixel) in 'super-resolution mode' (0.675 Å per pixel) at a dose rate of about nine electrons per pixel per second. Two movies were acquired per hole, and each movie encompassed a total dose of ~43 electrons per square ångström over 8 s fractionated into 40 frames (0.2 s each). Defocus values ranged from -0.6 µm to -3.1 µm. Movies were aligned and binned as previously described^{15,33}, except that images were not partitioned into quadrants.

Unless otherwise noted, processing was performed using RELION 1.3 (ref. 34). Contrast transfer function (CTF) parameters were estimated using CTFFIND4 (ref. 35). Initial 2D classes were generated after semi-automated picking of ~10,000 particles (box size 204) using e2boxer.py (EMAN2)³⁶. Sixteen distinct classes were low-pass filtered to 25 Å resolution and used as templates for autopicking³⁷, resulting in 476,100 particles selected from 1,172 micrographs. The autopicked particles were subjected to manual screening followed by screening by 2D classification, yielding an input data set of 409,401 particles. A previously published 22-Å-resolution cryo-negative stain reconstruction of human Pol II (EMD-1282)⁴ filtered to 50 Å was used as an initial reference for 3D refinement. Before any 3D classification, data were subjected to the particle polishing movie-processing algorithm of RELION 1.3 (ref. 38), resulting in an improvement in resolution from 3.7 Å to 3.4 Å.

Three-dimensional classification was performed without image alignment as outlined in Extended Data Fig. 2. Masks were chosen to include either the entire Pol II EC or a smaller region of interest. The full data set was used as input for the classification of heterogeneity in the region of upstream DNA density. Only classes displaying strong clamp density were used as input for classification of conformations of the RPB4-RPB7 stalk. After classification, data were again subjected to 3D refinement with a 50 Å filtered reference volume. B-factors were automatically estimated in RELION³⁹ and resolutions were reported on the basis of the gold-standard Fourier shell correlation (FSC) (0.143 criterion)⁴⁰ as described⁴¹. The Pol II EC1 reconstruction was calculated from 264,134 particles to 3.4 Å resolution and sharpened with a B-factor of -137 Å². The Pol II EC2 (improved RPB4-RPB7 stalk density) reconstruction was calculated from 219,265 particles to 3.6 Å and sharpened with a B-factor of -128 Å². The Pol II EC3 (improved upstream DNA density) reconstruction was calculated from 184,122 particles to 3.7 Å and sharpened with a B-factor of -123 Å².

Focused refinements were achieved by continuing a refinement of the full data set from the iteration at which local searches began, but replacing the mask encompassing the entire Pol II EC density with a soft mask around the region of interest and allowing the refinement to continue to convergence. Local resolution was

calculated using a sliding window method as described^{15,42}, except that a single pair of half maps was used per estimation and local resolution was not capped at the nominal value. Figures were generated using UCSF Chimera⁴³.

Model building and refinement. Alignments for each Pol II subunit were generated using the *Homo sapiens*, *Bos taurus*, *Drosophila melanogaster*, *Schizosaccharomyces pombe*, and *S. cerevisiae* sequences followed by alignment using Clustal Omega⁴⁴. The Pol II crystal structure PDB 4BBS⁴⁵ was chosen as a reference, as it displayed good stereochemistry and included the most complete model of the RPB2 protrusion domain. A starting model of ten-subunit Pol II was generated using the CCP4 (ref. 46) program chainsaw⁴⁷ along with the alignment of the *B. taurus* and *S. cerevisiae* sequences. Conserved amino acids were retained, and non-conserved amino acids were pruned back to the gamma atom. The starting model was placed in the Pol II EC density by fitting in UCSF Chimera⁴³, followed by fitting of rigid body groups in COOT⁴⁸. Groups for rigid body refinement were chosen on the basis of observations from Pol II X-ray studies and visual inspection of the initial fit to the density. In regions of sufficient density quality, the model was manually adjusted to include complete side chains, and missing or divergent regions were built in COOT using B-factor sharpened maps. Multiple maps were used during model building, including the densities for the refinement using all data, Pol II EC1, Pol II EC2, Pol II EC3, and focused refinements.

To generate a complete EC model, one molecule (chains A and B) of the crystal structure of human RPB4–RPB7 (PDB 2C35)⁴⁹ was docked into the Pol II EC2 map. Regions near the ten-subunit core were adjusted manually to fit the density. Amino-acid side chains (previously stubbed) were added to the model if side chain density was visible. Ideal B-form DNA was manually fitted into the Pol II EC3 upstream DNA density. To improve backbone geometry, the EC model was subjected to PHENIX real space refinement (global minimization and ADP refinement) into one of the three sharpened EC maps using Ramachandran, rotamer, and nucleic-acid restraints⁵⁰. EC3 was used for refinement of the upstream DNA (chain N residues 1–13 and chain T residues 27–39). Because of the lower local resolution of the distal end of the upstream DNA, residues 1–10 of the non-template strand and residues 29–39 of the template strand were replaced with ideal B-form DNA that had been aligned to the refined upstream nucleic acids. EC2 was used for refinement of RPB4 and RPB7, and EC1 was used for the remaining model (EC body). The EC body was additionally refined as described above using the sharpened EC1 map to better position well-resolved side chains and nucleic acids within the density⁵⁰. The final model was validated using Molprobity⁵¹, EMRinger⁵², and the FSC of the final model versus the EC1 map (Extended Data Fig. 3b). For model versus map FSC calculations, the EC1 map was masked using the RELION-generated soft automask used in postprocessing.

Crosslinking and mass spectrometry. Nucleic-acid scaffold used to assemble Pol II EC complexes was modified to include a 50 nucleotide RNA (sequence 5'-GAACGAGAUCAUAACAUAUUGAACAAGAUAUAUACAUAAGACCA GGC-3'), as previous data have shown that DSIF affinity for ECs is increased as RNA length increases⁵³. RNA was produced and purified as previously described⁵⁴. A twofold molar excess of pre-annealed RNA and template DNA was incubated with Pol II for 20 min at 25 °C, followed by incubation with fourfold excess of non-template DNA for an additional 20 min at 25 °C. A fivefold molar excess of DSIF was incubated with the resulting Pol II EC for 20 min at 25 °C. Pol II–DSIF EC sample was applied to two consecutive Superdex 200 10/300 size-exclusion columns equilibrated in buffer B (150 mM NaCl, 5 mM HEPES pH 7.25 at 25 °C, 10 μM ZnCl₂, 10 mM DTT). Purified Pol II–DSIF EC was concentrated to ~0.5 mg ml⁻¹ (~0.74 μM) and crosslinked with 3 mM BS3 (BS3-d0/d12, Creative Molecular) as described above. Crosslinked sample was again applied to two Superdex 200 10/300 columns, resulting in ~25 μg material. Sample was digested with trypsin, and analysed as previously described^{15,55}.

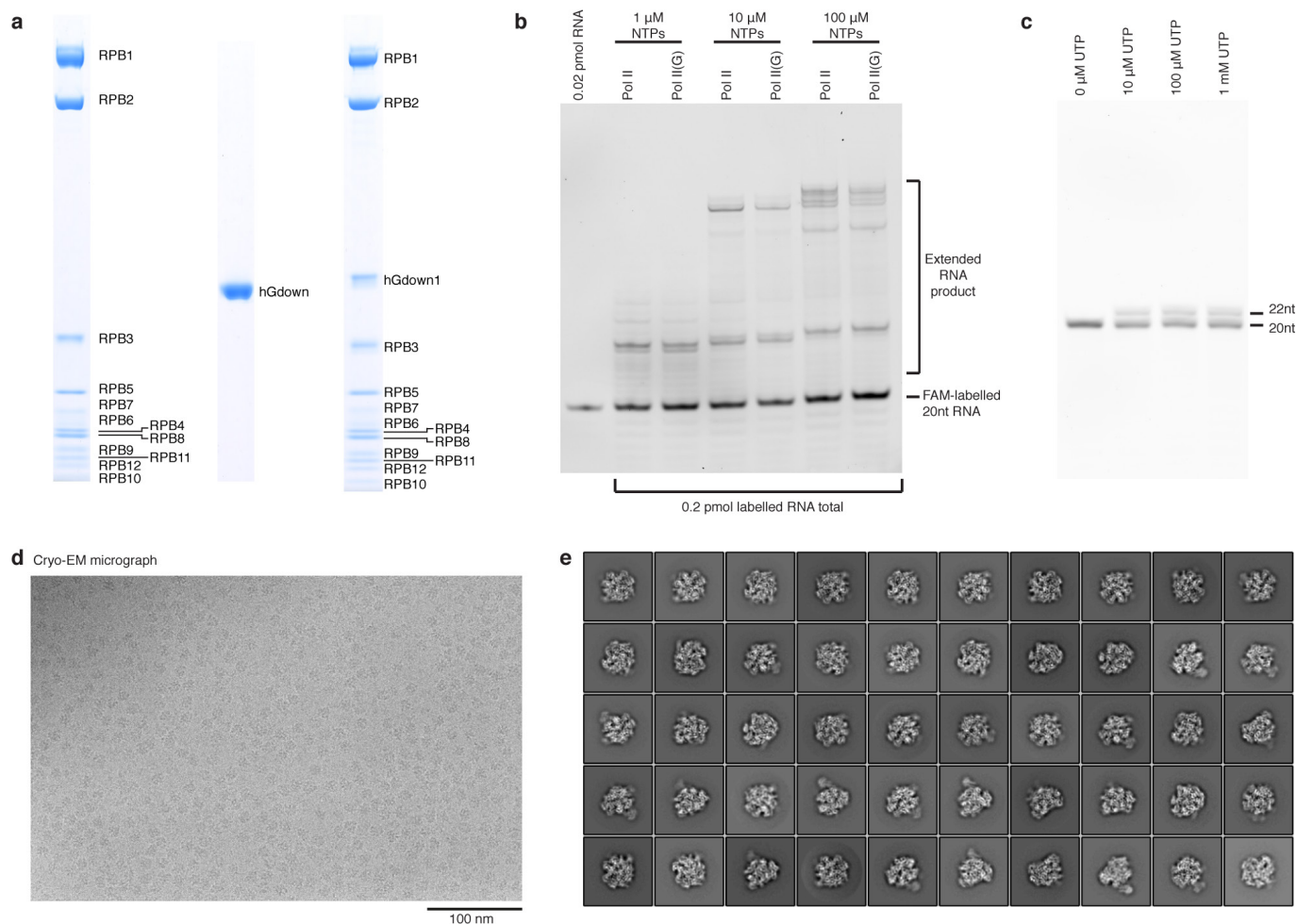
Negative stain electron microscopy of the Pol II–DSIF EC. Pol II–DSIF EC complexes were prepared as described above. Sample (~30 μg ml⁻¹) was applied to glow-discharged 400 mesh copper grids coated with continuous carbon (Plano EM) for 1 min, washed on 500 μl water for 30 s, floated for 20 s on three consecutive 20-μl drops of 2% uranyl formate stain, and blotted to remove excess stain. Data were collected using an FEI Tecnai Spirit operated at 120 kV and a magnification of ×90,600. Micrographs were collected using a defocus range from –1.0 to –1.5 μm with an FEI Eagle CCD (charge-coupled device) camera binned 2 × (image dimensions 2,048 pixels × 2,048 pixels) at a pixel size of 3.31 Å. Semi-automatic picking using e2boxer.py (EMAN2) yielded 11,531 particles from 120 micrographs. Data were subjected to 3D classification in RELION (eight classes, no CTF correction) using the cryo-negative stain reconstruction of human Pol II (EMD-1282)⁴ low-pass filtered to 60 Å as an initial reference. The two highest populated classes (comprising 85% of the data) were further classified into two classes each, for a total of four classes in which Pol II features beyond 60 Å were recognizable. Two classes did not have discernible additional density compared with Pol II (42% of data). A third class (28% of the data) displayed additional

density near the RPB4–RPB7 stalk. A final class (15% of the data) showed additional density over the Pol II DNA binding cleft, as well as additional density near the RPB4–RPB7 stalk. Refinement of this subset of data (1,630 particles) resulted in a 3D reconstruction at 26 Å resolution, revealing extra density consistent with results from crosslinking coupled to mass spectrometry, previous DSIF–RNA crosslinking⁵⁶, and the published interaction between SPT5 and the Pol II clamp coiled-coil motif²⁶.

Transcription assays. Activity assays were performed as described⁵⁷, with modifications. For reactions using fully complementary template and non-template DNA sequences, Pol II ECs were assembled stepwise beginning with either 12-subunit bovine Pol II or 12-subunit bovine Pol II in complex with human Gdown1, as indicated. Per reaction, Pol II was pre-assembled for 20 min at 28 °C with a 0.5 molar ratio of 5' 6-FAM-labelled 20-nucleotide RNA annealed to template DNA, followed by incubation with a 1.0 molar ratio of fully complementary non-template DNA. The RNA and DNA sequences were the same as for the Pol II EC, except for an additional 46 nucleotides of downstream DNA. The template DNA sequence was 5'-ACAAATTACTGGGAAGTCGACTATGCAATACAGGCAT CATTGTGATCAAGCTCAAGTACTTAAGCCTGGTCATTACTAGTACTGCC-3'; the non-template DNA sequence was 5'-GGCAGTACTAGTAATGACCAGG CTTAAGTACTTGAGCTTGATCAAATGATGCCTGTATTGCATAGTCGACT TCCCAGTAATTTGT-3', and RNA sequence was 5'-UAUAGCAUAAAGA CCAGGC-3'. Transcription was allowed to proceed for 10 min at 30 °C in the presence of 1–100 μM nucleoside triphosphates (NTPs) as indicated, and 0.2 pmol product per reaction was visualized on a 15% denaturing urea polyacrylamide gel. Transcription assays were also performed on the bubble scaffold used for structural studies. Pol II–EC complexes were prepared as described for cryo-EM, except that the samples were not crosslinked and the 20-nucleotide RNA used for assembly was 5'-labelled with 6-FAM. Assembled complex was incubated with 10–1,000 μM UTP at 30 °C for 10 min, allowing the extension of the RNA by two additional nucleotides. Product was visualized on a 20% denaturing urea polyacrylamide gel and imaged using a Typhoon FLA 9500.

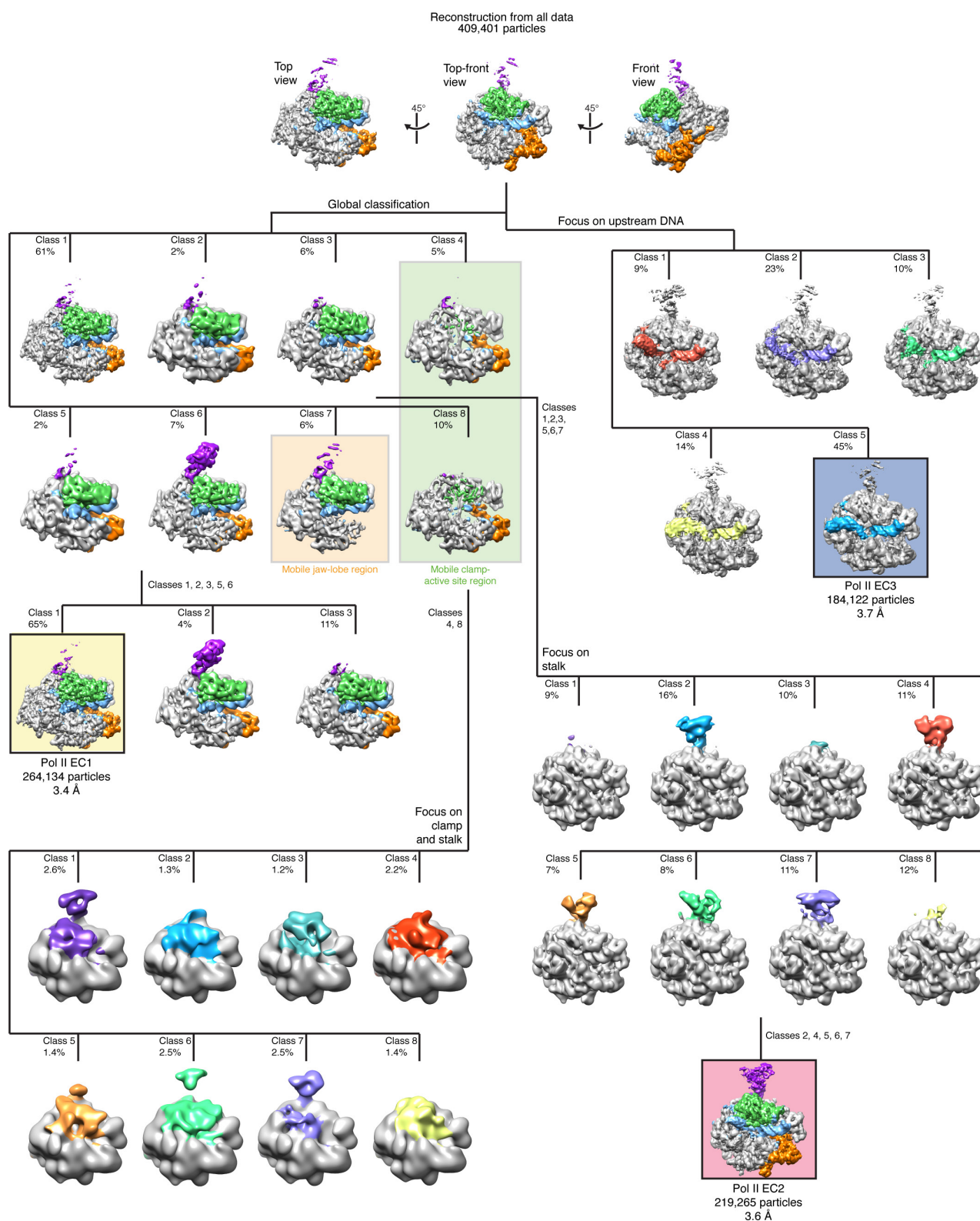
- Wenzel, S., Martins, B. M., Rösch, P. & Wöhr, B. M. Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochem. J.* **425**, 373–380 (2010).
- Korinek, A., Beck, F., Baumeister, W., Nickell, S. & Plitzko, J. M. Computer controlled cryo-electron microscopy—TOM² a software package for high-throughput applications. *J. Struct. Biol.* **175**, 394–405 (2011).
- Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
- Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
- Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
- Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
- Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
- Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
- Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
- Sainsbury, S., Niesser, J. & Cramer, P. Structure and function of the initially transcribing RNA polymerase II–TFIIB complex. *Nature* **493**, 437–440 (2013).
- Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
- Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643 (2008).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Meka, H., Werner, F., Cordell, S. C., Onesti, S. & Brick, P. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res.* **33**, 6435–6444 (2005).

50. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
51. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
52. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature Methods* **12**, 943–946 (2015).
53. Cheng, B. & Price, D. H. Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res.* **36**, e135 (2008).
54. Martinez-Rucobo, F. W. *et al.* Molecular basis of transcription-coupled pre-mRNA capping. *Mol. Cell* **58**, 1079–1089 (2015).
55. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348–1352 (2012).
56. Missra, A. & Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the *Drosophila* RNA polymerase II transcription elongation complex. *Proc. Natl Acad. Sci. USA* **107**, 11301–11306 (2010).
57. Brueckner, F., Hennecke, U., Carell, T. & Cramer, P. CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).
58. Brueckner, F. & Cramer, P. Structural basis of transcription inhibition by α -amanitin and implications for RNA polymerase II translocation. *Nature Struct. Mol. Biol.* **15**, 811–818 (2008).
59. Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).



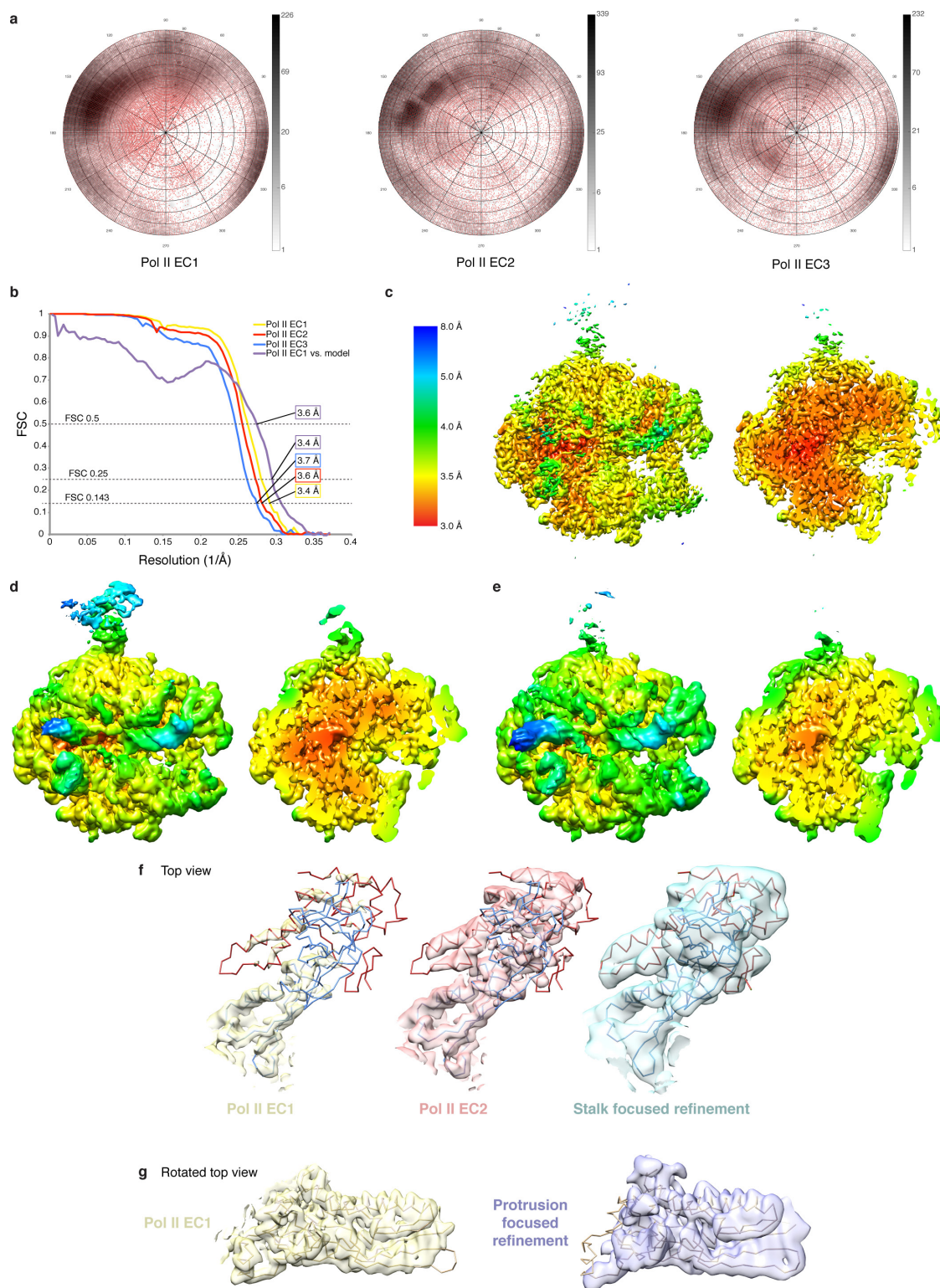
Extended Data Figure 1 | Bovine Pol II purification and transcription activity. **a**, SDS–polyacrylamide gel electrophoresis analysis of size-exclusion-purified 12-subunit bovine Pol II, recombinant Gdown1, and reconstituted complete Pol II with all 13 subunits (right). **b**, Extension of RNA by bovine Pol II. Activity assay was performed on Pol II ECs assembled on fully complementary DNA. Complexes were assembled with 12-subunit Pol II (Pol II) or 12-subunit Pol II in complex with Gdown1 (Pol II (G)). Transcription was allowed to proceed for 10 min at 30 °C in the presence of the indicated concentration of NTPs. Purified endogenous

bovine Pol II is active in transcription. Approximately 40% of the RNA appears to be extended in the presence of Gdown1. **c**, Extension of RNA by bovine Pol II–human Gdown1 on the bubble scaffold used for structural studies. Uncrosslinked Pol II EC was incubated with the indicated amount of UTP at 30 °C for 10 min to allow the incorporation of two additional nucleotides. Approximately 30% of RNA is extended. **d**, Observation of individual Pol II ECs of the expected size under cryo-EM conditions with a K2 direct electron detection device. **e**, Fifty representative 2D class averages resulting from classification of the full Pol II EC data set.



Extended Data Figure 2 | Computational sorting of cryo-EM particle images. Particle images were sorted by 3D classification in RELION to reveal additional features of the Pol II EC. To reduce computational costs, all classifications were performed without image alignment. For global classifications, a mask encompassing the entire Pol II EC was used, whereas for local classifications masks encompassed only the area of interest as indicated in the diagram. The percentage of total data included in each class is indicated above each panel. Density was coloured to assist in visualization. For global classifications (upper left), density was coloured according to observed mobile regions. Green: clamp, anchor, part of the RPB1 cleft, part of the hybrid-binding region, and part of the active

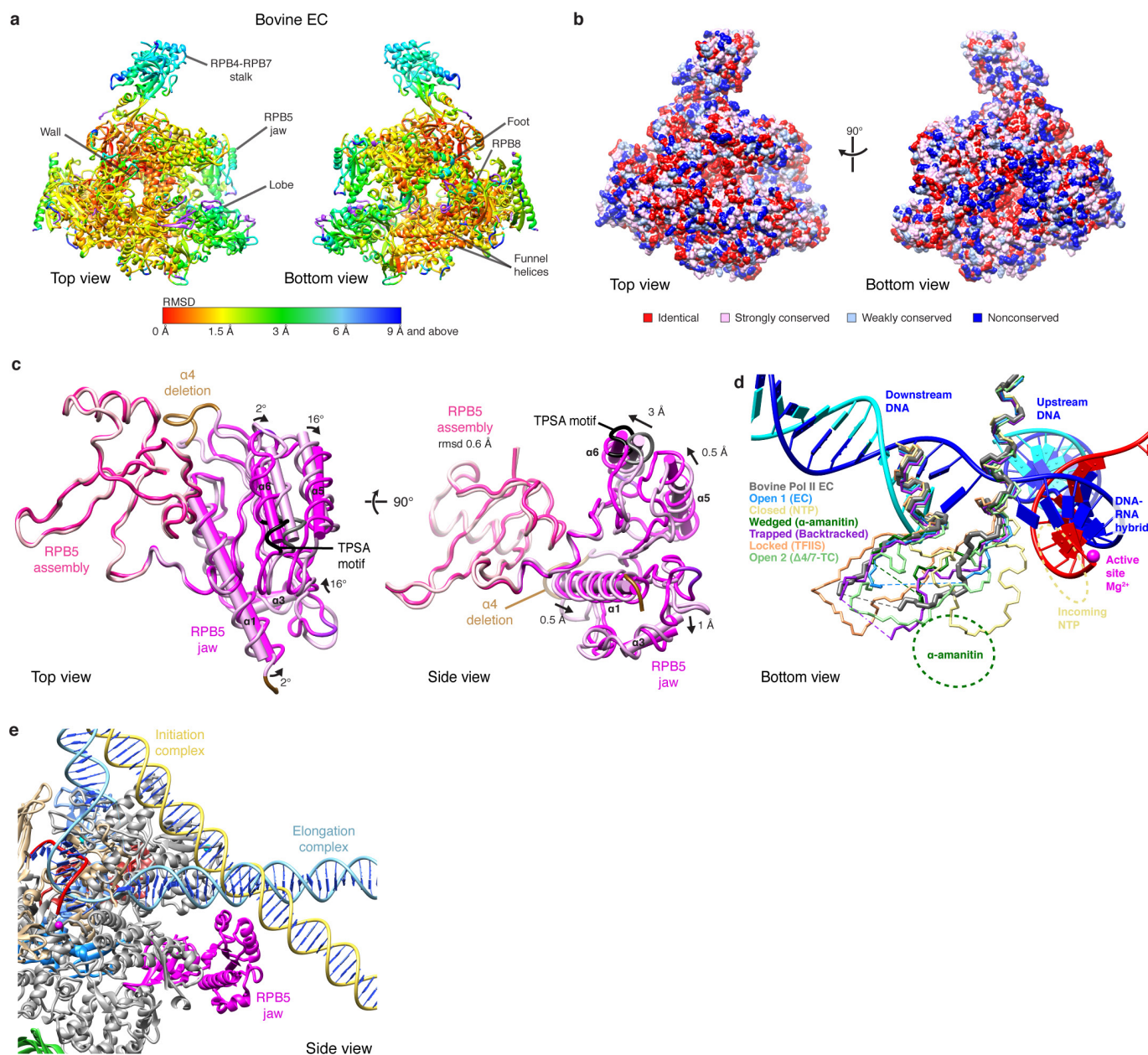
site. Orange: RPB9, RPB1 upper jaw, and part of the RPB1 cleft. Purple: RPB4–RPB7 stalk. Blue: nucleic acids. For focused classifications, density for the region of interest was coloured. This corresponded to the region within the classification mask, except for the classification of upstream DNA, in which case the density in the region of the entire nucleic-acid scaffold was coloured. The three reconstructions used for modelling the Pol II EC are shown in yellow, red, and blue boxes. The EC class displaying a mobile jaw-lobe region is highlighted with an orange box, and the EC class displaying a mobile clamp-active site region is highlighted with a green box.



Extended Data Figure 3 | Quality of the cryo-EM reconstructions.

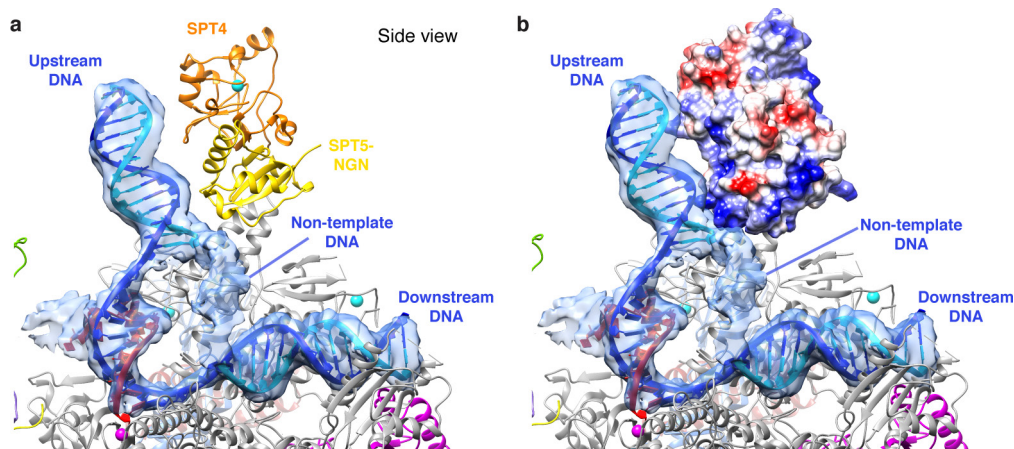
a, Angular distribution of particle images for the Pol II EC1, EC2, and EC3 reconstructions. Black shading indicates the number of particles assigned to a given view, while red dots indicate represented views (at least one particle was assigned within 1° of the point). **b**, FSC plots for the reconstructions described in **a**, as well as the FSC plot for the Pol II EC model versus the unsharpened Pol II EC1 map. The first data point after phase randomization is omitted. **c**, Top-front view and middle slice of the unsharpened Pol II EC2 reconstruction coloured by local resolution. The top-front view (also shown in Extended Data Figure 2) is a rotation of either the top view or front view by 45° . **d**, Top-front view and middle slice

of the unsharpened Pol II EC2 reconstruction coloured by local resolution. **e**, Top-front view and middle slice of the unsharpened Pol II EC3 reconstruction coloured by local resolution. **f**, RPB4-RPB7 stalk density for the unsharpened Pol II EC1, Pol II EC2, and stalk focused refinements shown at the same threshold level and filtered to the nominal resolutions (3.4 Å, 3.6 Å, and 4.2 Å). Continuous density for the stalk is visible in the lower resolution focused refinement. **g**, Density for the protrusion in the unsharpened Pol II EC1 and protrusion-focused refinements shown at the same threshold level and filtered to the nominal resolutions (3.4 Å and 3.8 Å). Density for the protrusion tip is more visible in the focused refinement map.



Extended Data Figure 4 | Structural comparison of bovine and yeast Pol II. **a**, Structures of the bovine and yeast EC (PDB 1Y1W) were superimposed by alignment of the active site. The protein regions of the bovine EC are shown in two views, coloured by r.m.s.d. of the yeast and bovine models. Insertions and uniquely structured regions within the bovine EC are coloured purple. **b**, Surface conservation between yeast and bovine Pol II. Strong and weak conservation groups were assigned according to Clustal conventions. **c**, Structural rearrangements within the RPB5 jaw domain. Axes were drawn through the four longest helices of the bovine and yeast RPB5 jaw using UCSF Chimera. **d**, Comparison of the backbone path of the trigger loop to previously known trigger loop conformations. The trigger loop backbone of the bovine Pol II EC is shown in thick, dark grey; the open conformation of the yeast EC (PDB 1Y1W)²⁰

is in blue; the closed conformation of the yeast EC in the presence of GTP (PDB 2E2H)²⁴ is in yellow; the wedged conformation of the yeast EC in the presence of α -amanitin (PDB 2VUM)⁵⁸ is in dark green; the trapped conformation of the arrested and backtracked yeast EC (PDB 3PO2)⁵⁹ is in purple; the locked conformation of the yeast EC reactivation intermediate in the presence of TFIIS (PDB 3PO3)⁵⁹ is in orange; and the new open position observed in the ten-subunit yeast transcription complex crystallized in the presence of TFIIF is in light green. The binding site for α -amanitin is denoted in dashed green lines. The binding site for the incoming NTP is outlined in dashed yellow lines. **e**, Comparison of the path of DNA in the closed initiation complex (yellow⁷) and the elongation complex (blue, this study). Downstream DNA was extended.



Extended Data Figure 5 | Modelled position of the human SPT5 NGN domain on the bovine Pol II EC. **a**, Zoom-in view of Pol II EC3 nucleic acid density superimposed on the Pol II-SPT4-SPT5-NGN EC model reveals proximity of non-template DNA to the SPT5 NGN domain. The human SPT4-SPT5-NGN crystal structure (PDB 3H7H)³¹ was positioned on the Pol II EC model by alignment with the archaeal RNA polymerase

clamp-SPT4-SPT5 crystal structure (PDB 3QQC)²⁶. **b**, Colouring a surface representation of the X-ray structure of the SPT4-SPT5-NGN by Coulombic surface charge reveals positively charged patches in close proximity to non-template and upstream DNA. Positive charge is in blue; negative charge is in red.

Extended Data Table 1 | Model refinement statistics

Map CC (whole unit cell)	0.835
rmsd (bonds)	0
rmsd (angles)	0.73
All-atom clashscore	14.17
Ramachandran plot	
outliers	0.00%
allowed	5.08%
favored	94.92%
Rotamer outliers	0.00%
C-beta deviations	0
EMRinger score	2.62
Molprobit score	2.01

Extended Data Table 2 | Summary of differences between the yeast EC (X-ray, PDB 1Y1W) and bovine EC (cryo-EM)

Subunit	Region	Sequence range B. taurus (S. cerevisiae)	Sequence change	Structural change
RPB1	Clamp core	36-84 (32-80)	no insertion or deletion	conformational change
	Clamp head	120-136 (116-130)	α 3- α 4 insertion	conformational change of α 3 and connecting loops
	Clamp head	156-180 (150-163)	α 4- β 4 insertion	loop expansion, conformational change
	Clamp head	203-212 (186-198)	β 4- β 5 deletion	conformational change
	Dock	431-444 (417-430)	no insertion or deletion	conformational change
	Pore 1	604-622 (590-599)	β 20- β 21 insertion	loop expansion, helix addition, conformational change
	Foot	908-946 (885-977)	no insertion or deletion	conformational change
	Cleft	1144-1150 (1121-1127)	no insertion or deletion	conformational change, helix addition
	Jaw	1171-1300 (1148-1275)	insertion and deletion (below)	domain-wide movement of loops and α -helices
	Jaw	1190-1205 (1167-1186)	α 40- β 29 deletion	loop shortening, α 40 conformational change
	Jaw	1262-1280 (1243-1255)	β 31- α 43 insertion	loop expansion
	Cleft	1312-1330 (1287-1300)	β 32- β 33 insertion	loop insertion, interruption of β 32 and β 33, conformational change
RPB2	Protrusion	68-81 (72-85)	no insertion or deletion	conformational change
	Protrusion	133-144 (137-162)	β 2- β 3 deletion	conformational change, β -strand addition
	Lobe	200-392 (218-405)	-	domain movement
	Lobe	209-215 (227-234)	β 7- β 8 deletion	loop shortening, conformational change
	Lobe	242-252 (261-265)	β 9- β 10 insertion	loop expansion
	Fork loop 1	453-466 (466-479)	no insertion or deletion	conformational change
	Fork loop 2	490-496 (503-509)	no insertion or deletion	conformational change
	External 1	630-632 (642-650)	β 21- α 16 deletion	loop shortening
	External 1	645-650 (663-678)	α 16- α 17 deletion	loop and α 16 shortening
	External 1	684-690 (712-735)	α 19- β 24 deletion	loop shortening
	Wall	820-827 (865-872)	no insertion or deletion	conformational change
	Wall	832-843 (877-888)	no insertion or deletion	conformational change
RPB3	Domain 2	123-147 (121-140)	β 7- β 8 insertion	loop expansion, conformational change
	Loop	197-202 (191-200)	β 10- β 11 deletion 1	loop shortening, conformational change
	Loop	209-222 (207-221)	β 10- β 11 deletion 2	conformational change
	Tail	264-271 (263-270)	no insertion or deletion	extension of α 3
RPB5	Jaw	45-52 (47-54)	α 3- β 1 insertion	conformational change
	Jaw	65-70 (65-76)	α 4 deletion	loop shortening, α 4 deletion
	Jaw	76-127 (82-132)	no insertion or deletion	conformational change
RPB8	β -barrel	16-20 (16-18)	β 2- β 3 insertion	loop expansion, β 2 shortening
	β -barrel	33-37 (31-36)	β 3- β 4 deletion	loop shortening, conformational change
	Loop	63-84 (62-89)	β 5- β 6 deletion	loop shortening, conformational change
	β -barrel	102-110 (107-111)	β 7- β 8 insertion	loop expansion
	β -barrel	129-138 (129-138)	β 9- β 10 insertion	conformational change
RPB9	Jaw	12-54 (2-44)	N-terminal insertion	domain movement
	Zinc ribbon 2	63-75 (53-65)	no insertion or deletion	conformational change
	Zinc ribbon 2	82-121 (72-108)	β 7- β 8 insertion	conformational change
RPB12	Zinc ribbon	15-16 (25-28)	N-terminal deletion	conformational change

CAREERS

HIV RESEARCH How a pathologist pursued an unconventional vaccine **p.557**

EMPLOYMENT Scientists launch project to ensure future jobs go.nature.com/uszpoj

NATUREJOBS For the latest career listings and advice www.naturejobs.com



ADAPTED FROM TOMACCO/GETTY

PSYCHOLOGY

Faking it

In the face of routine rejection, many scientists must learn to cope with the insidious beast that is impostor syndrome.

BY CHRIS WOOLSTON

Every early-career scientist has hit a stumbling block: a bad grade on an exam, a low score on a grant proposal or the first rejection from a journal. But for many, this normal stuff of science twists into something darker and more insidious — a creeping sense of professional inadequacy that prompts them to question their place in the field, no matter their

area of study or their level of brilliance.

There is a term for this type of self-doubt, coined in the 1970s by two US psychologists who saw it in their clinical practices: the 'impostor phenomenon'. Now commonly known as impostor syndrome, the condition can manifest itself in myriad professions, from office workers to artists to athletes, says Frederik Anseel, a psychologist at Ghent University in Belgium. Scientists, he says, are

especially vulnerable, largely because they work in a hero-oriented field that treats its highest achievers as if they were sports stars, leaving many others to wonder in silence whether they are second-stringers or worse. "Young people think that no one else is having these feelings," he says.

Researchers who struggle with the syndrome have to learn how to tune out feelings of inadequacy and develop a more realistic view of their abilities and their value, he says. In a profession where sporadic failure — in grants, in jobs, in publications — is the norm, the real failure is unnecessarily giving up on a promising career.

UNHAPPY SUCCESSES

In 2014, Anseel and his colleagues took a closer look at impostor syndrome in a study of more than 200 Belgian workers in finance, education and human-resource management. The team found that workers who reported feelings that are consistent with impostor syndrome tended to score higher on measures of neuroticism and excessive perfectionism in personality tests (J. Vergauwe *et al.* *J. Bus. Psychol.* **30**, 565–581; 2015). They were also not as happy with their jobs as were colleagues who did not experience the syndrome — even though some of the afflicted had advanced to the upper levels of their professions.

Anseel says that his other work — which includes ongoing studies of mental-health issues among young researchers — gives him confidence that his findings about impostor syndrome in the white-collar world apply to science as well. He says that it is easy to see how even successful scientists can feel that they are actually underperformers. Scientists, he says, often trivialize their own achievements. "You get a paper published in *PNAS*, and you tell yourself, 'That's doable. I'll never get a paper in *Nature* or *Science*.'" Similarly, any grant could be larger; any job could be better; any paper could be more highly cited. "You set yourself up to fail one way or the other," he says.

The phenomenon shows up across academia, including at top research institutions. Josh Drew, an evolutionary ecologist at Columbia University in New York City, has seen PhD and master's degree students struggle with self-doubt at the Ivy League school. Every student had passed tough admission standards — but that was not enough to bolster their confidence. For many, their classes at the university represented the first time in their educational experience that they didn't feel as if they were the smartest person in the ►

► room. “They were all outstanding students as undergrads,” he says. “Here, being at the top of your class is just average.”

In a highly competitive arena, self-doubt can be a career killer that prompts would-be contenders to dismiss chances to vie for important opportunities. “I saw many students who were shooting themselves in the foot,” Drew says. “They weren’t applying for grants and awards that they would be competitive for.” He began to address the syndrome in an introduction-to-graduate-school class. The talks drew some buzz, and he soon developed a formal presentation to deliver to other departments at Columbia and beyond (see ‘Help for impostors’). Clearly, he had struck a chord. “Every talk I give, people say, ‘I thought I was the only person who felt this way,’” he says.

TIMELESS CONDITION

Drew reassures people who feel like frauds by pointing out that they are in some lofty company. Two years after publishing *On the Origin of Species* in 1859, Charles Darwin complained that “one lives only to make blunders”. And while working on *The Grapes of Wrath* (1939), John Steinbeck wrote, “I am assailed by my own ignorance and inability,” fretting that “sometimes, I seem to do a good little piece of work, but when it is done it slides into mediocrity.”

While preparing his lecture, Drew solicited Twitter comments from scientists who had struggled to overcome the syndrome with various degrees of success. One respondent, an associate professor of biology, tweeted: “It has crippled my professional life from day one.” Moses Milazzo, a planetary scientist with the Astrogeology Science Center in Flagstaff, Arizona, tweeted, “Because of Impostor Syndrome: I have decided not to pursue opportunities; I am never ready to publish my papers; etc.”



Ecologist Josh Drew loads samples of reef fish into a tank of liquid nitrogen in Nagigi, Fiji.

SOCIAL SUPPORT

Help for impostors

Evolutionary ecologist Josh Drew of Columbia University in New York City hosts a talk called ‘Fighting the Impostor Syndrome’, which offers tried-and-tested coping strategies for the common condition. At a basic level, he urges researchers to advocate for themselves. That means avoiding words such as ‘just’ and ‘only’ when describing their own work, and not constantly apologizing for every mistake, whether real or perceived. He says that offering real support to someone else who feels wracked by doubt is a quick and

effective way to improve your own sense of belonging. To really pick yourself up, bring someone else up with you.

Drew says that members of groups that are underrepresented in science often benefit by reaching out to others and finding a community. For some, simply following a Twitter feed such as #BLACKandSTEM or #womenandSTEM can serve as reassurance that they really do belong in science. His message? “You’re not here because you ticked some box. You’re here because you bring a lot to the department.” **C.W.**

Milazzo thinks that impostor syndrome is nearly universal among scientists — at least among those who are self-aware enough to realize that they don’t know everything. But he also says that he experienced particularly severe effects of it, and he traces at least some of his unease to his background. He grew up in an off-the-grid adobe house on the border of the Navajo Nation reservation in northern Arizona, which later contributed to his sense that he did not belong in the university crowd. Not many of his instructors, advisers or peers could relate to hardscrabble desert life in a house that had no reliable electricity and little contact with the outside world.

“I didn’t even know what *Cosmos* was until I got to grad school,” he says of the popular 1980s TV show. Then again, life in middle-of-nowhere Arizona did give him an intimate familiarity with the night sky. During the long walk along a dirt road to the school bus stop, he often navigated by the light of the Milky Way.

Milazzo says that he first heard the term impostor syndrome early in his graduate-student days at the University of Arizona in Tucson — and he recognized it immediately. “Having a name put to it made it clear that other people felt it,” he says. But knowing that he was not alone didn’t keep him out of the trap. He decided not to apply for a NASA grant for satellite-based research, out of fear of exposing his own ignorance. “I removed myself from the grant process because it would be obvious that I didn’t know what I was talking about,” he says. As it turned out, one of the successful proposals was very similar to his idea. “If I had pursued it, I might have been competitive,” he laments now.

And even after helping to win a NASA grant to develop a middle-school curriculum that would be based on the space agency’s exploration of the Solar System, Milazzo still struggles to persuade himself that he belongs in science. “We put a lot of work into that proposal, but I wasn’t very confident that it would get funded,” he says.

Matt von Hippel, a researcher at the Perimeter Institute for Theoretical Physics in Waterloo, Canada, says that he, too, feels like an impostor from time to time, but he has a strategy that helps him to push through it. Instead of second-guessing the people who admitted him to graduate school and awarded him a PhD, he decided to embrace their judgement.

“You can trust the system to have put you in vaguely the right job,” he says. “If you’re invited to give a talk, that’s a sign that you’re ready to give a talk.” Late last year, he was asked to give a colloquium on mathematical techniques in particle physics at Oregon State University in Corvallis. It was a big opportunity that came sooner in his career than he expected, and he thought about turning it down. Ultimately, he opted to adhere to his strategy. “I decided to say ‘yes’ and see how it goes,” he says. In his view, the talk was a success.

UPHILL BATTLE

Biologist Victoria Metcalf had plenty of opportunities to doubt herself and second-guess her career choices. Her low point came in early 2000 during her PhD studies in New Zealand, when television-news crews surrounded her lab and a regulatory authority threatened to throw her and her supervisors in prison. Her lab had cloned genes from the tuatara (*Sphenodon punctatus*) — a treasured native New Zealand reptile — but lacked the permits that a new law had retroactively made mandatory. Authorities eventually dropped their threats, but her research was stalled for six months while she obtained the proper permits. “Those were really soul-destroying times,” she says. “It had a huge impact on how I perceived my worth in academia.”

Scientists are accustomed to measuring things in precise detail, but their own value can be difficult for them to quantify. Ansell thinks that many researchers would be more confident — and thus more likely to write the grant, submit the paper, apply for the job — if they were to embrace the inevitability of

failure. “When one of my students gets a rejection letter, I can show them five or ten of my own,” he says. “The academic environment should be more open to failure stories.”

Drew reminds young researchers that even the chairs of their departments — scientists who seemingly have it made — do not always get their grants funded or their papers accepted. It would be telling, he says, if everyone published a ‘shadow CV’ of all their rejections to go along with the standard CV that lists successes.

Researchers can also help to ease their distress by making an effort to stop comparing themselves with colleagues in their lab or department. “Comparisons won’t make you happy, so don’t do it,” Anseel says. Instead, he says, researchers should set their own personal standards of achievement and then do their best to meet them.

Metcalf has mostly won her battle over her sense of inadequacy, although her career has had its ups and downs. After she earned her PhD, she took a postdoc position in the United States that she quit after only six months,

an outcome that made her feel even more like a scientific impostor. “I had a low sense of self-worth,” she says. But she pushed through it, quickly found another post and

went on to have a successful career that included research trips to the Antarctic and a highly sought-after faculty position at Lincoln University in Christchurch, New Zealand.

Yet her troubles didn’t end. In 2011, she lost her faculty job after an earthquake damaged much of the city. Instead of taking that setback as a sign that she needed to abandon science completely, she shifted from research to outreach. She is now the national coordinator of the Participatory Science Platform, a New Zealand government programme that promotes research collaborations between scientists and communities. “Anyone who knows me knows that I was meant for this job,” she says.

As part of her duties, Metcalf has had many chances to speak to young people with different backgrounds and career aspirations. Many of them are already experiencing the symptoms of impostor syndrome, which gives her an opportunity to inspire by example. “My story really resonates,” she says. “I’ve had my battles. You just have to keep fighting.” ■

Chris Woolston is a freelance writer in Billings, Montana.

TURNING POINT

Louis Picker

Louis Picker is not afraid to break with convention. Trained as a pathologist, he was on the front line when the AIDS epidemic emerged in the 1980s. He is now combining his interests in immunology and viruses to pursue an unusual HIV vaccine at Oregon Health and Science University (OHSU) in Portland — a project that was considered a fool’s errand by many when he began.

How did you get started in research?

I had always wanted to be a scientist. I started an MD–PhD programme at the University of California, San Francisco, but found it much too slow, rigid and hierarchical. I left that programme, but did a year of research there. Ultimately, I decided to become a pathologist specializing in immunology. It’s astonishing how much biology you can learn from looking at hundreds of biopsy slides and by performing autopsies every day. I got a feel for the immune system that you couldn’t get by doing graduate research on a mouse.

Describe your first AIDS autopsy.

I was a pathology resident at Beth Israel Hospital in Boston, Massachusetts. The devastation left by AIDS stuck with me. I decided to learn more about the disease so that I could do something about it one day. I had the opportunity to move into HIV research in the mid-1990s and haven’t looked back since.

What led you to HIV-vaccine research?

Early in my career, I worked on a flow-cytometry-based assay to measure specific T-cell responses to viral infection in humans. I chose to work with cytomegalovirus (CMV), a virus that infects around 50% of adults in the United States and triggers a T-cell response that lasts throughout a person’s lifetime. These factors enabled me to test the specificity of the assay. After studying CMV-specific T cells, I hypothesized that CMV could be exploited to create a vaccine that stimulates an immediate immune response to a variety of pathogens. By incorporating bits of HIV into the vaccine, we could prime T cells to hit the intruding virus early and hard. Our data in non-human primate models show that the vaccine stops infection with the simian counterpart of HIV in slightly more than half of recipients.

What does the next year hold for you?

We will move into clinical trials with our potential HIV vaccine. We are also exploring the use of unconventional viral vectors to manipulate the immune system against



tuberculosis, malaria, hepatitis B and cancer at a level heretofore unappreciated.

Why did you choose research over more-lucrative private practice?

I knew that if I wanted to make a difference — and to pursue the CMV-based vaccine while others focused on conventional antibody-led approaches — I had to do lab-based experiments. As a pathologist, I would never have had access to patients. The best way to do relevant science was to test my ideas in a non-human primate model. The job I took at the OHSU was one of two possibilities I had at the time to do that type of work.

How easy was it to pursue your idea?

I was fortunate to have negotiated a start-up package at the OHSU that gave me the leeway to gamble. Either I’d make it or break it. I was warmly welcomed by researchers in the HIV field, which I appreciated. But it took me a while to feel that I fit in. Self-doubt was a powerful driver for me.

How risky was your decision?

To be honest, it helped that I had an MD. I knew I would always be able to get a job as a physician, so the degree allowed me a little more freedom in the early years. In the first crucial years while I was establishing myself, I figured I could always return to pathology. Most people with PhDs don’t have that option.

What makes a great scientist?

You have to be a little bit of a lunatic. But your out-of-the-box thinking also has to be right. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

ROBOT BURIAL

A moment to reflect.

BY H. E. ROULO

Sanders stared across the scrubby hillside, resting a hand on his shovel. The horizon wavered in the heat. He sighed. Maybe he shouldn't have answered when Richie texted, but they had been chums in school. Anyway, it was too late to back out now. Sweat trickled behind his ears and dripped onto the stiff collar of his white dress shirt. He swiped at the drops.

The crunch of Richie's shovel stilled and he looked up from digging. Hazy grey dust filled the brackets around his mouth. He had hardly spoken since Sanders had arrived at their old bonfire spot, except to point to the extra shovel leaning against the twisted pine tree. A tree that, by the way, cast almost no shade.

He noticed Richie watching him with a scowl and frowned back. "What?"

Richie licked his lips and finally spoke. "Are you making fun of me in that get-up? A suit?"

"No, I just heard the word burial, and it was instinctive." Sanders closed his eyes and took a deep breath. The whole thing certainly had the grimness of a funeral. He raised the wooden shaft of his shovel out of the hole they stood in and tapped the robot's padded frame, hearing the metal clang. "I've never buried a robot before."

Richie's face turned red, like he'd finally realized how he had sounded when he called and asked Sanders to meet him in the middle of nowhere for a rush burial. Who could blame Sanders for being rattled? What a relief to confirm the body really was a powered-off jumble of foam, plastic and metal.

Richie jumped out of the hole and grabbed a sports bottle, taking a long pull of blue sugar-water. He snapped the bottle's cap back into place and nearly set the bottle on the frame of the robot but fumbled, placing it on the ground instead.

"That's a change from what we used to drink out here." Relieved they'd stopped digging, Sanders leaned on his shovel, wondering if Richie was ready to talk. Even when they were in school together he'd kept



his thoughts private, and Sanders was the one who filled the silences. Some things never change. "You can still see bottles piled at the base of the tree. Looks like kids don't come out here anymore, though."

Richie examined the tree, perhaps thinking of girls he had kissed under its branches, back when they were struggling students looking for a night away from it all. He mumbled: "Those were good times."

Not like today, seemed to go unsaid.

"Maybe the suit was silly," Sanders admitted. "It's not like I expected we'd meet at a cemetery. What graveyard is going to allow a robot to be buried there? And it's not like the robot was religious, am I right?"

Richie chuckled.

Encouraged, Sanders said: "She was a household robot, right? They're different now, but I remember one like that from when I was little."

"My folks were always early adopters. The robot was bought to assist my mother with chores even before her diagnosis."

"Gotcha. Old story. I suppose she ended up being a nanny to you? You spent more time with her than your parents?"

Richie's lips tightened.

"Even with true AI outlawed, it's easy to form attachments."

Folks always did for pets, why not for robots? Especially if they talk?" Sanders said. "I bet we end up with robot cemeteries like they have pet cemeteries. We should invest in one. We'd make a quick buck for sure."

Richie shook his head. "No, she wasn't for me. As Mom's muscles grew weaker, the robot could be trained to take over her burdens. Pretty soon, she was taking care of Mom, too. As Mom got older, the robot was the only one who knew exactly what she wanted. Maybe I should have been there more, but I knew she'd be taken care of."

Uncertain of where this was going, Sanders leaned over to examine the robot. "She's dinged up, but are you sure she's dead?"

Richie shook himself like a dog. "She's a collection of Mom's quirks, and now that Mom's dead, she doesn't have a purpose. Too old to retrain, and too beat-up to sell. I called several dealers. They said they would buy the parts." With his left hand, he reached out to the prone robot and folded flesh-toned padding into place. "But after everything she's done, it would be like selling my sister for parts. She didn't just do the things my parents didn't want to do, she did what I didn't want to do, too. Mom and I both got to keep nearly normal lives for a long time after she was disabled ... I completed my degree because of her."

"I kept her running for a while, but I don't want her to be the part of my mother that lasts the longest. It's time."

Sanders was nodding. "So, we bury the robot."

"Yeah," Richie picked up his shovel. "But I'll have to run back to my house for something."

Reluctantly, Sanders picked up his shovel. "What's that?"

"My good suit." ■

H. E. Roulo lives in Seattle. She has released a dozen short stories, including 'Immeasurable' in *Nature* previously. She recently released the first book in her *Plague Master* series. Find more at www.heroulo.com.

➔ **NATURE.COM**

Follow Futures:

🐦 @NatureFutures

🌐 go.nature.com/mtoodm